

01-27-00

A

01/25/00

UTILITY PATENT APPLICATION TRANSMITTAL <small>(Only for new nonprovisional applications under 37 CFR 1.53(b))</small>	Title of Invention	Mosaic Protein and Restriction Endonuclease Assisted Ligation Method for Making the Same
	Named Inventor(s)	Yury E. Khudyakov Howard A. Fields
	Attorney Docket	03063-0381
	Express Mail Label No.	EL498679563US

APPLICATION ELEMENTS		Assistant Commissioner for Patents ADDRESS TO: Box Patent Application Washington, D.C. 20231	
1. <input checked="" type="checkbox"/> Fee Transmittal Form <i>(Submit an original, and a duplicate for fee processing)</i> 2. <input checked="" type="checkbox"/> Specification, Claims, and Abstract 3. <input checked="" type="checkbox"/> Drawings 4. Oath or Declaration a. <input type="checkbox"/> Newly executed (original or copy) b. <input checked="" type="checkbox"/> Copy from prior application (37 CFR 1.63(d)) <i>(for continuation/divisional with Box 17 completed)</i> [Note Box 5 Below] (i) <input type="checkbox"/> <u>DELETION OF INVENTOR(S)</u> Signed statement attached deleting inventor(s) named in the prior application, see 37 CFR 1.63(d)(2) and 1.33(b). 5. <input checked="" type="checkbox"/> Incorporation by Reference <i>(usable if Box 4b is checked)</i> The entire disclosure of the prior application, from which a copy of the oath or declaration is supplied under Box 4b, is considered as being part of the disclosure of the accompanying application and is hereby incorporated by reference therein. 6. <input type="checkbox"/> Microfiche Computer Program (Appendix) 7. <input checked="" type="checkbox"/> Nucleotide and/or Amino Acid Sequence Submission (if applicable, all necessary) a. <input checked="" type="checkbox"/> Request to Use Computer Readable Form of Sequence Listing from Another Application Under 37 CFR 1.821(e) b. <input type="checkbox"/> Paper Copy (identical to computer copy) c. <input type="checkbox"/> Statement verifying identity of above copies		ACCOMPANYING APPLICATION PARTS 8. <input checked="" type="checkbox"/> Parent Application assigned to Centers for Disease Control and Prevention 9. <input type="checkbox"/> 37 CFR 3.73(b) Statement <i>(when there is an assignee)</i> <input type="checkbox"/> Power of Attorney by assignee 10. <input type="checkbox"/> English Translation Document (if applicable) 11. <input checked="" type="checkbox"/> Information Disclosure Statement (IDS) PTO-1449 <input type="checkbox"/> Copies of IDS Citations 12. <input checked="" type="checkbox"/> Preliminary Amendment 13. <input checked="" type="checkbox"/> Return Receipt Postcard (MPEP 503) <i>(Should be specifically itemized)</i> 14. <input type="checkbox"/> Small Entity Statement(s) <input type="checkbox"/> Statement filed in prior application Status still proper and desired 15. <input type="checkbox"/> Certified Copy of Priority Document(s) 16. <input checked="" type="checkbox"/> Other: <u>25 sheets formal drawings</u> <hr/> <hr/>	
17. If a CONTINUING APPLICATION , check appropriate box and supply the requisite information: <input type="checkbox"/> Continuation <input checked="" type="checkbox"/> Divisional <input type="checkbox"/> Continuation-in-part (CIP) of prior application No: 08/921,887			
18. CORRESPONDENCE ADDRESS: Jamie L. Greene, Esq. JONES & ASKEW, LLP 2400 Monarch Tower 3424 Peachtree Road, N.E. Atlanta, Georgia 30326			
		By: <u>Jamie L. Greene</u> Reg. No. 32,467 Date: <u>Jan. 25, 2000</u> Telephone: <u>404-949-2400</u> Facsimile: <u>404-949-2499</u>	

00491146 012500

01/25/00

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of:

YURY E. KHUDYAKOV AND
HOWARD A FIELDS

Filed: Concurrently Herewith

For: MOSAIC PROTEIN AND RESTRICTION
ENDONUCLEASE ASSISTED LIGATION
METHOD FOR MAKING THE SAME

Art Unit: 1643

Examiner: _____

PRELIMINARY AMENDMENTAssistant Commissioner for Patents
Washington, DC 20231

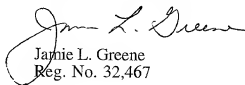
Sir:

Prior to examination of the above-referenced patent application, please enter the following amendments.

In the Claims

Please cancel Claims 1-12 and 20-30.

Respectfully submitted,



Jamie L. Greene
Reg. No. 32,467

JONES & ASKEW, LLP
2400 Monarch Tower
3424 Peachtree Road, N.E.
Atlanta, Georgia 30326
(404) 949-2400
Our Docket: 03063-0381

Express Mail Mailing Label No. EL498679563US

Date of Deposit: Jan 25, 2000

05210 "STATION"

**MOSAIC PROTEIN AND
RESTRICTION ENDONUCLEASE ASSISTED
LIGATION METHOD FOR MAKING THE SAME**

Field of the Invention

The present invention relates to the field of recombinant production of proteins from synthetic genes. In particular, the invention relates to the expression of mosaic proteins constructed from antigenic peptides of the hepatitis C virus.

Background of the Invention

The hepatitis C virus (HCV) is classified as part of the *Flaviviridae* family, and contains a single, positive strand of RNA approximately 9400 nucleotides long, encoding for at least a 3000 amino acid polyprotein, depending on the source of the viral isolate. (Choo Q-L., *et al.*, *Proc. Natl. Acad. Sci. USA* 88:2451-2455, 1991; Choo Q-L., *et al.*, *Science* 244:359-362, 1989; Kato N., *et al.*, *Proc. Natl. Acad. Sci. USA* 87:9524-9528, 1990; Takamizawa A., *et al.*, *J. Virol.* 65:1105-1113, 1991) The 5'-end of the genome encodes for the structural proteins that include the nucleocapsid protein (C), and two envelope proteins (E1, and E2/NS1), whereas the 3'-end of the genome encodes for the non-structural proteins that include the NS2, NS3, NS4, and NS5 proteins. (Miller R.H. and Purcell R.H., *Proc. Natl. Acad. Sci. USA* 87:2057-2061, 1990; Takeuchi K., *et al.*, *J. Gen. Virol.*

71:3027-3033, 1990; Choo Q-L., *et al.*, *Proc. Natl. Acad. Sci. USA* 88:2451-2455, 1991; Hijikata M., *et al.*, *Proc. Natl. Acad. Sci. USA* 88:5547-5551, 1991; Takamizawa A., *et al.*, *J. Virol.* 65:1105-1113, 1991; Houghton M., *et al.*, *Hepatology* 14:381-388, 1991) The hepatitis C virus is the main causative agent of non-A, non-B hepatitis, and plays a major role in the development of chronic liver disease, liver cirrhosis, and hepatocellular carcinoma worldwide. (Choo Q-L., *et al.*, *Proc. Natl. Acad. Sci. USA* 88:2451-2455, 1991) Since there is no vaccine and no effective therapy for HCV induced disease, diagnosis and prevention of infection are issues of major public health importance.

In an effort to improve the efficiency of HCV diagnosis, many antigenic regions have been identified along the HCV genome, and used to develop three generations of enzyme immunoassays for the detection of anti-HCV activity in human sera. (Kuo *et al.*, *Science* 244:362-364, 1989; Chien D.Y., *et al.*, *Proc. Natl. Acad. Sci. USA* 89:10011-10015, 1992) Each successive generation represented an improvement in both the sensitivity and specificity of the enzyme immunoassay (EIA) by adding more antigenic regions to the assay. The first generation of enzyme immunoassays relied on the detection of antibodies to a region within a non-structural protein, 5-1-1. Second and third generation assays were based on the detection of antibodies against the recombinant 22 kDa core or nucleocapsid (NC) protein and several recombinant proteins derived from non-structural regions of the viral polyprotein (NS3, NS4, NS5).

Although the improvements in the specificity and sensitivity of the EIAs have resulted in a reduction in the number of new HCV infections (Alter H.J., *Blood* 85(7):1681-1695, 1995), many investigators have indicated that the current versions still require further development. (Tobler L.H., *et al.*, *Transfusion* 34:130-134, 1994; Courouce A.M., *et al.*, *Transfusion* 34:790-795, 1994; Damen M., *et al.*, *Transfusion* 35:745-749, 1995; Feucht H.H., *et al.*, *J. Med. Virol.* 48:184-190, 1995; Bar-Shany S., *et al.*, *Inetnrl. J. Epi.* 25:674-677, 1996; Dhaliwal S.K., *et al.*, *J. Med. Virol.* 48:184-190, 1996; Pawlotsky J.M., *et al.*, *J. Clin. Micro.* Jan:80-83, 1996) The impetus to improve tests for detection of

anti-HCV is based upon studies demonstrating that currently available EIAs have relatively poor specificities, especially in low-prevalence populations. (Alter H.J., *Blood* 85(7):1681-1695, 1995; Feucht H.H., *et al.*, *J. Med. Virol.* 48:184-190, 1995) Additionally, even after the development of supplemental tests, such as MATRIX (Abbott Laboratories, Abbott Park, IL), used to confirm EIA positive sera, 10% of specimens are still classified as indeterminate (reactive to a single antigen) following supplemental testing. (Pawlotsky J.M., *et al.*, *J. Clin. Micro.* Jan:80-83, 1996) These findings might be due to testing sera during the very early stage of infection before all antibodies reach detectable levels. Alternatively, reactivity to a single antigen may be due to non-specificity of the specimen.

Another important limitation to currently available assays is the use of genotype 1 recombinant proteins as immunologic targets. Recently, it was reported that there are differences in the serologic reactivity of the current EIAs to the different HCV genotypes. (Zein N.N., *et al.*, *Mayo Clinic Proc.* 70(5):449-452, 1995; Dhaliwal S.K., *et al.*, *J. Med. Virol.* 48:184-190, 1996) This observation suggests that the current EIAs, which are based on type I HCV, may need to be further improved by including antigenic epitopes from different genotypes. Therefore, it is clear that there remains a strong need for an enzyme immunoassay with increased specificity and sensitivity that would react with sera infected with multiple genotypes of the hepatitis C virus.

To create an enzyme immunoassay with broad reactivity to multiple genotypes, a synthetic protein must be assembled from a long DNA fragment containing multiple antigenic epitopes. The synthesis of long artificial DNA polynucleotides has been made possible by the availability of highly efficient methods to chemically synthesize relatively short oligonucleotides. To assemble a gene from oligonucleotides, several enzymatic reactions using polymerases and/or ligases may be used. Two methods described elsewhere (Khudyakov Y.E., *et al.*, *Nucleic Acid Res.* 21(11):2747-2754, 1993; Khudyakov Y.E., *et al.*, *J. Virol.* 68:(11)7067-7074, 1994; and U.S. Patent No.

5,563,032), the polymerase chain reaction (PCR) and the Exchangeable Template Reaction (ETR), have been successfully applied to assemble synthetic genes from oligonucleotides. The use of PCR, however, is disadvantageous in cases where repeated sequences are designed in the gene, while ETR can not be used to conveniently express short fragments of the synthetic gene. Therefore, a new method of assembling synthetic genes with repeated sequences that would allow for the expression of shorter fragments of the gene, would greatly facilitate the creation of a synthetic protein to be used in an improved enzyme immunoassay, in particular for the detection of HCV.

Summary of the Invention

The present invention relates to methods and compositions used to improve the sensitivity, the spectrum of immunoreactivity, and the specificity of antigens used as immunologic targets for detection. In preferred embodiments, the detection can be performed by enzyme immunoassay (EIA). The method, designated Restriction Endonuclease Assisted Ligation (REAL), involves the construction of an artificial gene from synthetic oligonucleotides. The compositions are synthetic proteins composed of a mosaic of broadly immunoreactive antigenic epitopes from several genotypes of, for example HCV.

REAL employs the use of the Klenow fragment of DNA Polymerase I to convert specially designed complementary oligonucleotides into double stranded DNA fragments, which are subsequently amplified by PCR. Restriction sites were engineered into the cloning vector and used to produce complimentary overhangs for the addition of consecutive fragments. Each fragment may be cloned and expressed individually, for example in *Escherichia coli* to determine their immunoreactivity or may be assembled into full length product without cloning. Two consecutive fragments are subsequently ligated, amplified by PCR, cleaved with restriction endonucleases, and ligated with DNA ligase to assemble each fragment into a longer fragment in a consecutive process. By repeating this process fragments of increasing length are assembled, expressed and analyzed for

immunoreactivity, and reiterated until the full length gene is assembled.

The invention provides mosaic proteins comprising a plurality of homologous antigenic peptides from different genotypes of a hepatitis virus. In particular, the invention provides mosaic proteins comprising a plurality of homologous antigenic nucleocapsid peptides from different genotypes of a hepatitis C virus. Further, the invention provides mosaic proteins comprising a plurality of homologous antigenic non-structural peptides from different genotypes of a hepatitis C virus. The mosaic proteins and genes encoding therefor can be used for immunologic detection or vaccination.

Brief Description of the Figures

Figure 1 is a schematic representation of the strategy used to insert each monomer into the cloning vector. Monomers were cloned into EcoRI and BamHI treated vector pCV3 by treatment of annealed oligonucleotides with the Klenow fragment of DNA Polymerase I followed by restriction endonuclease cleavage.

Figure 2 is a schematic representation of the pCV3 vector design, and demonstrates that the pGEX-4T-2 vector (SEQ ID NO:53) was treated with BamHI and NotI to remove the multiple cloning site. The recovered cleaved vector was then treated with Klenow fragment of DNA polymerase I in the presence of GTP and ATP to prevent self annealing of plasmid.

Figure 3 shows the ligation of a new multiple cloning site sequence (SEQ ID NO:54) into cloning vector by DNA ligase to create the vector pCV3 (SEQ ID NO:55).

Figure 4 is a schematic representation of the design of the cloned segment. The bar represents a specific immunoreactive epitope from an appropriate etiologic agent.

Figure 5 depicts the cloning strategy of each monomer. The fragments were treated with EcoRI and BamHI followed by ligation into similarly treated pCV3.

Figure 6 is a schematic representation of the PCR amplicons containing BbvI, EcoRI, BamHI, and FokI restriction

sites, and the subsequent restriction of the first monomer with the FokI restriction enzyme.

Figure 7 shows the restriction of a second consecutive monomer with the BbvI restriction enzyme.

Figure 8 demonstrates the ligation of a FokI treated first monomer with a BbvI treated second monomer with DNA ligase to create a dimer.

Figure 9 shows the amino acid sequence of each monomer comprising the NC mosaic protein. The bolded amino acids indicate the variations between the sequences of each monomer.

Figure 10 depicts the strategy of assembly from individual NC monomers by REAL resulting in a full length gene composed of monomers ABCDEFGHIJK. The primary sequence of all cloned fragments were confirmed prior to assembly by REAL.

Figure 11 shows the analysis of selected expressed and purified protein by immunoblot analysis. A single anti-HCV positive human sera with high anti-NC activity by MATRIX was used as the source of primary antibody.

Figure 12 shows the frequency distribution of 200 anti-HCV negative sera by NC Mosaic EIA for the detection of anti-NC activity. The cutoff value was statistically derived as the mean optical density (OD) value of negative sera plus 3.5 times the standard deviation of the mean. This value equals 0.145.

Figure 13 shows an endpoint titration of specimen no. BBI 304 by NC Mosaic EIA and by MATRIX for the detection of anti-NC activity.

Figure 14 shows an endpoint titration of specimen no. BB1325 by NC Mosaic EIA and by MATRIX for the detection of anti-NC activity.

Figure 15 depicts a comparison of the NC Mosaic EIA (OD y-axis) and MATRIX (S/C y-axis) for the detection of anti-NC activity in seroconversion no. 4812.

Figure 16 depicts a comparison of the NC Mosaic EIA (OD y-axis) and MATRIX (S/C y-axis) for the detection of anti-NC activity in seroconversion no. 4813.

Figure 17 depicts a comparison of the NC Mosaic EIA (OD y-axis) and MATRIX (S/C y-axis) for the detection of anti-NC activity in seroconversion no. 4814.

Figure 18 shows the amino acid sequence of the NS4 mosaic protein. A total of 17 sequences were used corresponding from the same domains from different subtypes and genotypes. Domain 1691 - 1710 corresponds to the N-terminus of 5-1-1, while domain 1712 - 1733 corresponds to the C-terminus of 5-1-1. Domain 1921 - 1940 corresponds to region 59 (C-terminus of NS4).

Figure 19 demonstrates the strategy of NS4 gene assembly by Restriction Endonuclease Assisted Ligation (REAL). The mosaic NS4 gene was assembled from 9 consecutive monomers (A, B, C, D, E, F, G, H, and I). The monomers were first assembled into 4 dimers (AB, CD, EF, and GH), and then 4 dimers were further assembled into 2 tetramers (ABCD and EFGH). The full length gene was constructed by the ligation of 2 tetramers and 1 monomer (I).

Figure 20 shows the analysis of expressed and affinity-chromatography purified proteins of all monomers (A, B, C, D, E, F, G, H and I), dimers (AB, CD, EG, and GH), tetramers (ABCD, and EFGH), and the full length gene (A - I) by immunoblot using a single anti-NS4 positive specimen diluted 1:200. Three monomers (A, H, and I) were not immunoreactive even when the specimen was diluted 1:10.

Figure 21 depicts the frequency distribution of 160 anti-HCV negative sera and 166 anti-HCV positive sera by NS4 Mosaic EIA.

Figures 22A-B shows the endpoint titration of specimen no. 1 (Figure 22A) and no. 2 (Figure 22B) by the NS4 Mosaic EIA and by MATRIX.

Figure 23 shows a comparison of the percent positively for anti-NS3, anti-NS4, and anti-NC activity by MATRIX versus anti-NS4 activity by NS4 Mosaic EIA using 182 anti-HCV positive sera.

Figures 24A-D demonstrate the sensitivity of the NS4 Mosaic EIA versus MATRIX for the detection of anti-NS4 activity

in seroconversion panel nos. 4811 (Figure 24A), 4812 (Figure 24B), 6214 (Figure 24C) and 4813 (Figure 24D).

Figure 25 shows the reactivity of genotype specific sera by NS4 Mosaic EIA for anti-NS4 activity.

Detailed Description

The present invention relates to methods and compositions used to improve the sensitivity, the spectrum of immunoreactivity, and the specificity of antigens used as immunologic targets for detection. In preferred embodiments, the detection can be performed by enzyme immunoassay (EIA). The method, designated Restriction Endonuclease Assisted Ligation (REAL), involves the construction of an artificial gene from synthetic oligonucleotides. The compositions are synthetic proteins composed of a mosaic of broadly immunoreactive antigenic epitopes from several genotypes of a species, for example hepatitis C virus (HCV). The mosaic protein compositions can be used for immunologic detection of, or vaccination against, the organisms from which they are derived. The invention further contemplates that the nucleic acids encoding the mosaic proteins can be used for immunologic detection of, or vaccination against, the organisms from which they are derived. Therefore, in addition to compositions and methods for detecting hepatitis, the invention provides a hepatitis vaccine comprising a mosaic protein, or a gene encoding therefor.

REAL employs the use of the Klenow fragment of DNA Polymerase I to convert specially designed complimentary oligonucleotides into double stranded DNA fragments, which are subsequently amplified by PCR. Restriction sites were engineered into the cloning vector and used to produce complimentary overhangs for the addition of consecutive fragments. Each fragment may be cloned and expressed individually, for example in *Escherichia coli* to determine their immunoreactivity or may be assembled into full length product without cloning. Two consecutive fragments are subsequently ligated, amplified by PCR, cleaved with restriction endonucleases, and ligated with DNA ligase to assemble each fragment into a longer fragment in a

consecutive process. By repeating this process fragments of increasing length are assembled, expressed and analyzed for immunoreactivity, and reiterated until the full length gene is assembled.

5 In particular, the present invention provides a method of constructing an artificial gene, comprising synthesizing an initial oligonucleotide containing an initial gene segment encoding an initial gene product. The initial gene segment is flanked in the upstream direction (5') by an upstream initial ligating sequence, a
10 first endonuclease recognition sequence that is recognized by a first endonuclease that cleaves at the first endonuclease recognition sequence, and a second endonuclease recognition sequence which is recognized by a second endonuclease that cleaves downstream of the first endonuclease recognition sequence and within the upstream initial ligating sequence.

15 Additionally, the initial gene segment is flanked in the downstream direction (3') by a downstream initial ligating sequence, a stop codon, a third endonuclease recognition sequence that is recognized by a third endonuclease that cleaves at the third endonuclease recognition sequence, and a fourth endonuclease recognition sequence which is recognized by a fourth endonuclease that cleaves upstream of the third endonuclease recognition sequence, upstream of the stop codon, and within the downstream initial ligating sequence. An example of such an initial
20 oligonucleotide can be seen in Figure 4, wherein "SEGMENT" designates the initial gene segment and "NNNNNN" designates the initial ligating sequence.

25 The method further comprises synthesizing a subsequent oligonucleotide containing a subsequent gene segment encoding a subsequent gene product. The subsequent gene segment is flanked in the upstream direction (5') by an upstream subsequent ligating sequence, a first endonuclease recognition sequence which is recognized by the first endonuclease that cleaves at the first endonuclease recognition sequence, and a second endonuclease recognition sequence which is recognized by the second
30 endonuclease that cleaves downstream of the first endonuclease
35

recognition sequence and within the upstream subsequent ligating sequence.

5 The subsequent gene segment is flanked in the downstream direction (3') by a downstream subsequent ligating sequence, a stop codon, a third endonuclease recognition sequence which is recognized by the third endonuclease that cleaves at the third endonuclease recognition sequence, and a fourth endonuclease recognition sequence which is recognized by the fourth endonuclease that cleaves upstream of the third endonuclease recognition sequence, upstream of the stop codon, and within the downstream subsequent ligating sequence. An example of such an subsequent oligonucleotide can be seen in Figure 4, wherein "SEGMENT" designates the subsequent gene segment and "NNNNNN" designates the subsequent ligating sequence.

10 The method further comprises the step of cleaving the initial oligonucleotide with the fourth endonuclease, and cleaving the subsequent oligonucleotide with the second endonuclease. An example of this step is shown in Figures 6 and 7.

15 The method further comprises the step of ligating the initial oligonucleotide and the subsequent oligonucleotide together at the downstream initial ligating sequence of the initial oligonucleotide and the upstream subsequent ligating sequence of the subsequent oligonucleotide to form an artificial gene. An example of this step is shown in Figure 8. The invention contemplated that additional subsequent oligonucleotides can be prepared, cleaved and ligated in a likewise fashion to make any artificial gene.

20 In a preferred embodiment, the invention provides the subsequent step of cleaving the artificial gene with the first and third endonucleases and inserting the remaining artificial gene into a vector previously cleaved with the first and third endonucleases. This step permits insertion of the final vector construct into a living organism, such as *E. coli*, and the recombinant expression of the artificial gene to produce the mosaic protein.

25 In a preferred embodiment, the invention provides the subsequent step of cleaving the artificial gene with the first and third endonucleases and inserting the remaining artificial gene into a vector previously cleaved with the first and third endonucleases. This step permits insertion of the final vector construct into a living organism, such as *E. coli*, and the recombinant expression of the artificial gene to produce the mosaic protein.

30 The invention provides the unique opportunity following the synthesizing the initial oligonucleotide step, and

35

before the cleaving the initial oligonucleotide with the fourth endonuclease step, of confirming the operability of the initial oligonucleotide by cleaving the initial oligonucleotide with the first and third endonucleases and inserting the remaining initial oligonucleotide into a vector previously cleaved with the first and third endonucleases, and expressing the initial gene segment. An example of such an additional step is shown in Figure 5, wherein "SEGMENT" designates the initial gene segment and "NNNNNN" designates the initial ligating sequence. This step is made possible by the inclusion of the stop codon downstream of the gene segment, which is removed by the addition of the fourth endonuclease in subsequent steps.

In preferred embodiments, the first and third endonucleases are EcoRI and BamHI, respectively. However, it will be understood that any endonucleases that cleave at the recognition sequence, can be used, with the proviso that two different endonucleases are employed. Examples of other suitable restriction endonucleases include: AflIII, Alw44I, ApaI, ApaII, BclII, BglIII, BspHI, BssHII, HindIII, KpnI, MluI, NarI, NcoI, PstI, SalI, or XhoI.

In preferred embodiments, the second and fourth endonucleases are BbvI and FokI, respectively. However, it will be understood that any endonuclease that cleaves downstream of the first endonuclease recognition sequence, or upstream of the third endonuclease recognition sequence, respectively, and within the ligating sequences, can be used, with the proviso that two different endonucleases are employed. Examples of other suitable restriction endonucleases that restrict the nucleic acid at a site away from the recognition site include: BspMI, HgaI, MboII, or SfaNI.

The invention provides that in preferred embodiments the initial and subsequent gene segments encode antigenic regions of a homologous protein from different genotypes of a hepatitis virus. The invention contemplates, however, that the REAL technique can be used for the construction of any mosaic or chimeric protein. In preferred embodiments, the gene segments encode antigenic regions of homologous proteins of different

genotypes of a hepatitis C virus. Preferably, the gene segments encode antigenic regions of a nucleocapsid protein or a non-structural protein of different genotypes of a hepatitis C virus.

5 The invention further provides an artificial gene constructed by the above methods. The invention further provides a mosaic protein encoded by the artificial gene constructed by the above methods. The invention further provides a method of detecting a hepatitis infection in an individual comprising combining a serum sample from the individual with the mosaic protein made by the above methods, and detecting the presence of antibody binding to the mosaic protein, the presence of binding indicating a hepatitis infection in the individual. Preferably, an enzyme immunoassay (EIA) is performed for detection of a hepatitis infection, as is described in Examples 2 and 3. The detection of antibody binding can be facilitated by the use of detectable moieties, such as fluorescence, radioisotopes or solid substrate capture.

10 The invention provides mosaic proteins comprising a plurality of homologous antigenic peptides from different genotypes of a hepatitis virus. In particular, the invention provides mosaic proteins comprising a plurality of homologous antigenic nucleocapsid peptides from different genotypes of a hepatitis C virus. Further, the invention provides mosaic proteins comprising a plurality of homologous antigenic non-structural peptides from different genotypes of a hepatitis C virus. The invention also provides the gene sequences which encode for such mosaic proteins.

15 In one preferred embodiment, the mosaic protein can comprise a plurality of homologous antigenic nucleocapsid peptides from different genotypes of a hepatitis C virus as set forth in the amino acid sequences set forth in SEQ ID NOs:23-33, detailed in Example 2, herein. In another preferred embodiment, the mosaic protein can comprise a plurality of homologous antigenic non-structural peptides from different genotypes of a hepatitis C virus as set forth in the amino acid sequence of SEQ ID NO:52, detailed in Example 3, herein. It will be understood that certain minor or silent amino acid modifications and/or

substitutions can be made in the amino acid sequences, while maintaining the antigenic functionality of the mosaic proteins. It will also be understood that certain silent or wobble nucleotide modifications and/or substitutions can be made in the gene sequences, while maintaining the ability of the gene to be ligated by the REAL technique and the antigenic functionality of the encoded mosaic proteins.

This invention is further illustrated by the following examples, which are not to be construed in any way as imposing limitations upon the scope thereof. On the contrary, it is to be clearly understood that resort may be had to various other embodiments, modifications, and equivalents thereof which, after reading the description herein, may suggest themselves to those skilled in the art without departing from the spirit of the present invention and/or the scope of the appended claims. The references mentioned in this specification are hereby incorporated by reference in their entireties.

Example 1

Restriction Endonuclease Assisted Ligation

Each monomer, or DNA segment encoding a peptide of interest, was prepared from synthetic oligonucleotides of 40-80 nucleotides that overlapped at 8-10 nucleotides and contained an EcoRI site on the 5'-terminus of the plus strand and a BamHI site on the 5'-terminus of the negative strand. More detailed examples of such monomers can be seen in SEQ ID NOs:1-22 and 34-51 described in Examples 2 and 3 herein. Single strand DNA was converted to double stranded DNA by adding 50 pmol of each oligodeoxynucleotide, 2 μ l BamHI buffer (Boehringer Mannheim, Indianapolis, IN) and 2.5 pmol dNTPs in a total volume of 20 μ l. The mixture was heated to 95°C for 1 minute and cooled on ice. After adding 1 μ l of the Klenow fragment of DNA polymerase I (Promega, Madison, WI), the reaction mixture was incubated 30 minutes at room temperature. An incubation step at 65°C for 10 minutes was used to inactivate the DNA polymerase. The resulting double stranded DNA molecule was cleaved with the restriction enzymes EcoRI and BamHI and ligated into pCV3 using the T4

DNA fast ligation kit (Boehringer Mannheim, Indianapolis, IN). (Figure 1) All internal BbvI, EcoRI, BamHI, and FokI restriction endonuclease sites were previously removed from the synthetic gene by introducing synonymous modifications into codons.

5 The pCV3 vector was created by modification of the multiple cloning site in the expression vector pGEX-4T-2 (Pharmacia, Piscataway, NJ). After cleaving the vector with BamHI and NotI to remove the internal EcoRI site, the recovered vector was treated with the Klenow fragment of DNA polymerase I (Promega, Madison, WI) in the presence of GTP and ATP to prevent self annealing of the restricted plasmid. (Figure 2) Subsequently, a double-stranded DNA molecule was prepared by annealing two complimentary oligonucleotides containing one EcoRI site flanked by a BbvI site, and one BamHI site flanked by a FokI site. This DNA molecule was inserted into the multiple cloning site of the modified vector by DNA ligase. (Figure 3) The final structure of pCV3 was confirmed by restriction enzyme analysis and by DNA sequencing.

10 After each monomer was inserted into the pCV3 vector, it was amplified by PCR using plasmid specific primers thereby acquiring BbvI and FokI sites and each structure was confirmed by restriction endonuclease analysis and DNA sequencing. Figure 4 shows the elements for each cloned monomer or segment of the synthetic gene. The solid line separating each strand represents the coding sequence of each monomer followed by a stop codon so that each fragment may be individually expressed. The EcoRI and BamHI sites are used for cloning, while the BbvI and FokI sites are used to remove the EcoRI and BamHI sites and to produce overhangs complimentary to the next consecutive monomer.

25 The process of consecutive assembly of monomers or fragments into a synthetic gene is illustrated in Figures 5 - 8. The first monomer is restricted with EcoRI and BamHI and ligated into similarly treated pCV3. Following amplification by PCR each segment acquired the restriction sites BbvI and FokI (Figure 4). Restriction with BbvI creates a 5' overhang 8 base pairs downstream of a GCAGC sequence, whereas restriction with FokI

35

creates a 5' overhang 9 base pairs downstream of a GGATG sequence. Therefore, upon reversing the order of the FokI site on the anti-sense strand of the DNA to create a 3' overhang, and situating the restriction sites 8 and 9 base pairs, respectively, from 4-nucleotide complimentary sequences, each consecutive monomer was alternately treated with FokI and BbvI to produce complimentary overhangs, a 3'-overhang on the first segment and a 5'-overhang on the second segment (Figures 6 and 7). The restricted segments were separated by agarose gel electrophoresis using 25% low melting point agarose and stained with ethidium bromide. After staining with ethidium bromide and cutting out the bands with correct size, the DNA was extracted by using the Wizard PCR preps (Promega, Madison, WI) following the instructions of the manufacturer. Bands corresponding to the restricted segments were recovered by melting the agarose, and each segment was purified using the Wizard PCR Preps DNA Purification System (Promega, Madison, WI).

Subsequently, the first amplified segment was treated with FokI to produce a 5'-overhang complimentary to the 3'-overhang on the second monomer created by treatment with BbvI. Each of the treated segments were purified and ligated with DNA ligase. When two monomers were treated as described above, a dimer was formed which could be subsequently treated with EcoRI and BamHI, and cloned into similarly treated pCV3. Since each segment was supplied with its own stop codon, each monomer, dimer, and multimer could be expressed and analyzed individually.

Example 2

Design and production of an artificial nucleocapsid mosaic protein

Analysis of data derived from Genebank

The N-terminal region spanning amino acids 5 - 33 of the nucleocapsid (NC) protein was selected as the region to develop a mosaic protein because of the presence of several strong and broadly immunoreactive antigenic epitopes (unpublished data). All known HCV sequences deposited in Genebank were analyzed using the program FASTA (Wisconsin Package 9.0, Genetics

Computer Group (GCG), Madison, WI). Out of 77 available sequences demonstrating variations in this region, 11 variants representing 3 different HCV genotypes were selected for designing an artificial mosaic gene. Figure 9 represents the amino acid sequence of each of the 11 variants, referred to as fragments or monomers, comprising the mosaic NC protein.

Synthetic gene assembly

The construction of the artificial gene was accomplished using the previously described Restriction Endonuclease Assisted Ligation (REAL) method and the following synthetic oligonucleotides:

A1: 5' - CCC CGA ATT CAA CCG AAA CCG CAA CGT AAA ACC AAA CGT AAC ACC ATT CGT CGT C (SEQ ID NO:1);

A2: 5' - CCC CGG ATC CTA TTT CGG ACC AAC GAT CTG ACC ACC ACC CGG GAA TTT AAC GTC CTG CGG ACG ACG AAT (SEQ ID NO:2);

B1: 5' - CCC CGA ATT CAA CCG AAA CCG CAA CGT CAG ACC AAA CGT AAC ACC AAC CGT CGT (SEQ ID NO:3);

B2: 5' - CCC CGG ATC CTA TTT CGG ACC AAC GAT CTG ACC ACC ACC CGG GAA TTT AAC GTC CTG CGG ACG ACG GTT G (SEQ ID NO:4);

C1: 5' - CCC CGA ATT CAA CCG AAA CCG CAA CGT AAA ACC AAA CGT AAC ACC TAC CGT (SEQ ID NO:5);

C2: 5' - CCC CGG ATC CTA TTT CGG ACC AAC GAT CTG ACC ACC ACC CGG GAA TTT AAC GTC CTG CGG ACG ACG GTA GGT G (SEQ ID NO:6);

D1: 5' - CCC CGA ATT CAA CCG AAA CCG CAA CGT AAA CCG AAC CGT AAC ACC AAC CGT CGT C (SEQ ID NO:7);

D2: 5' - CCC CGG ATC CTA TTT CGG ACC AAC GAT CTG
ACC ACC ACC CGG GAA TTT AAC GTC CTG CGG ACG
ACG GTT (SEQ ID NO:8);

5 E1: 5' - CCC CGA ATT CAA CCG AAA CCG CAA CGT CAG
CCG AAA CGT AAC ACC CCG CGT CGT CCG CAG GAC
(SEQ ID NO:9);

10 E2: 5' - CCC CGG ATC CTA TTT CGG ACC AAC GAT CTG
ACC ACC ACC CGG GAA TTT AAC GTC CTG CGG (SEQ ID
NO:10);

15 F1: 5' - CCC CGA ATT CAA CCG AAA CCG CAA CGT AAA
ACC AAA CGT AAC GCT CAC CGT CGT C (SEQ ID NO:11);

20 F2: 5' - CCC CGG ATC CTA TTT CGG ACC AAC GAT CTG
ACC ACC ACC CGG GAA TTT AAC GTC CTG CGG ACG
ACG GTG SEQ ID NO:12);

25 G1: 5' - CCC CGA ATT CAA CCG AAA CCG CAA CGT AAA
AAC CAG CGT AAC ACC AAC CGT CGT C (SEQ ID NO:13);

G2: 5' - CCC CGG ATC CTA TTT CGG ACC AAC GAT CTG
ACC ACC ACC CGG GAA TTT AAC GTC CTG CGG ACG
ACG GTT (SEQ ID NO:14);

H1: 5' - CCC CGA ATT CAA CCG AAA CCG CAA CGT AAA
ACC AAA CGT AAC ACC ATt CGT CGT C (SEQ ID NO:15);

30 H2: 5' - CCC CGG ATC CTA TTT CGG AAC GTA GAT AAC
ACC ACC ACC CGG GAA TTT AAC GTC CTG CGG ACG
ACG AAT (SEQ ID NO:16);

35 I1: 5' - CCC CGA ATT CAA CCG AAA CCG CAA CGT AAA
ACC GAA CGT AAC ACC AAC CGT CGT CC (SEQ ID NO:17);

I2: 5' - CCC CGG ATC CTA TTT CGG ACC AAC GAT CTG
ACC ACC ACC AGA GAA ACG AAC GTC CGG ACG ACG GT
(SEQ ID NO:18);

J1: 5' - CCC CGA ATT CAA CCG AAA CCG AAA CGT CAG
ACC AAA CGT AAC ACC CTG CGT CGT (SEQ ID NO:19);

J2: 5' - CCC CGG ATC CTA TTT CGG ACC AAC GAT CTG
ACC ACC ACG CGG GAA TTT AAC GTT TTT CGG ACG
ACG ACG CAG G (SEQ ID NO:20);

K1: 5' - CCC CGA ATT CAA CCG AAA CCG CAA CGT AAA
ACC AAA CGT AAA GCT CAC CGT CGT C (SEQ ID NO:21);

K2: 5' - CCC CGG ATC CTA TTT CGG ACC AAC GAT CTG
ACC ACC ACC CGG GAA TTT AAC GTC CTG CGG ACG
ACG GTG (SEQ ID NO:22).

Each pair of oligonucleotides was converted into double stranded DNA by the Klenow fragment of DNA Polymerase I and subsequently cloned, resulting in 11 monomers of 28 amino acids. Prior to sequential assembly by REAL, each fragment was amplified by the polymerase chain reaction, and determined to be the expected molecular weight by agarose gel electrophoresis with ethidium bromide staining. Additionally, the primary structure of each fragment was confirmed by sequencing. To begin assembly of the mosaic protein, two consecutive monomers were assembled into 5 dimers as shown in Figure 10. In the next step, the remaining monomer, G, was fused with the dimer EF to form a trimer, while the other consecutive dimers were assembled into the tetramers ABCD and HIJK. Fragments EFG and HIJK were then assembled into a septamer, and the septamer was assembled into a full length, 924 base pair gene by adding the tetramer ABCD.

Protein expression and purification

Proteins were fused to the C-terminus of glutathione S-transferase by transforming competent *Escherichia coli* cells, JM 109 (Invitrogen, San Diego, CA), with plasmids containing each of the fragments. Cells were grown in LB medium containing 100 μ g Ampicillin per ml in a bacteria shaker at 37°C until the optical density at 600 nm was equal to 0.6. The tac promoter was activated to achieve protein expression by adding isopropyl-b-D-thiogalactoside (IPTG) at a final concentration of 1 mM. After 1 hour growth at 30°C, the cells were harvested, and a lysate was prepared following the procedure described by Sambrook J., *et al.*, in Molecular Cloning - A Laboratory manual, latest edition., p. 17.38, Cold Spring Harbor Laboratory Press, New York, 1989. The glutathione S-transferase-mosaic fusion proteins were then purified by affinity chromatography using glutathione-Sepharose columns (Pharmacia, Piscataway, NJ) (Smith D.B. and Johnson K.S., *Gene* 67:37-40, 1988).

Analysis of NC expressed fragments

E. coli cells were transformed with plasmid constructs containing each of the PCR amplified fragments. After induction with IPTG, crude lysates were prepared and high yields of proteins of the expected molecular mass were observed after analysis by 12% SDS-PAGE (data not shown). A comparison of different induction conditions indicated that induction with 1 mM IPTG for 1 hour at 30°C gave the highest yield of soluble mosaic-fusion proteins (data not shown). Following the preparation of lysates, the proportion of soluble protein was estimated to be about 50 - 60%. Each expressed NC fragment was purified by affinity chromatography according to the manufacturer's recommendations, and analyzed by 12% SDS-PAGE and Coomassie blue staining. All 21 purified proteins demonstrated a high degree of purity and electrophoresed to their expected molecular weights. Although an artifactual doublet was present in many of the samples, this result is typical of the glutathione S-transferase (GST) expression system. The full length NC mosaic

protein electrophoresed as a single band with an estimated molecular weight of 61 kDa.

Immunoblot assay

To verify the immunoreactivity of each fragment, the GST-mosaic fusion proteins were analyzed by immunoblot using an anti-HCV positive sample having high anti-NC activity by MATRIX. Nitrocellulose membranes containing immobilized proteins were incubated for 1 hour with anti-NC positive human sera diluted 1:200 times in washing solution (0.1M PBS, pH 7.2, containing 1% BSA, and 0.5% Tween 20). The membranes were washed three times with washing solution and then incubated for 1 hour with affinity-purified goat anti-human immunoglobulin G conjugated to horseradish peroxidase (Biorad, Richmond, CA) diluted 1:5000 in washing solution. After washing, diaminobenzidine and hydrogen peroxide were added to develop the color reaction. As shown in Figure 11 (asterisks indicate the location of specific immunoreactivity), each of the purified proteins demonstrated immunoreactivity suggesting the accessibility of immunoreactive epitopes. The monomers were the least immunoreactive, and as the fragments increased in size they became increasingly more immunoreactive. Many of the lanes corresponding to the higher molecular weight fragments demonstrate specific reactivity to proteolytic cleavage products. Although Figure 11 shows data for 16 of the 21 proteins, the remaining 5 proteins behaved in a similar manner.

NC Mosaic EIA

Twenty nanograms of full length affinity-purified GST-mosaic NC fusion protein in PBS (pH 7.5) was added to microtiter wells (Immunolon II: Dynatech Laboratories, Inc., Chantilly, VA) and allowed to adsorb at room temperature for 12 hours after which the wells were blocked with 10% normal goat serum (NGS), and 1% BSA in PBS for 2 hours at 37°C. Human sera diluted 1:500 in 0.1 M phosphate-buffered saline, pH 7.5, containing 0.1% Tween 20 and 10% NGS was added and incubated for 1 hour at 37°C. After washing, goat anti-human IgG

conjugated to horseradish peroxidase diluted 1:5000 in 0.1 M PBS, pH 7.5, containing 0.1% Tween 20 and 10% NGS was added, and the wells were incubated for 1 hour at 37°C. The wells were incubated for ten minutes in the dark with substrate. Acid was added to stop the reaction and optical density (OD) was measured at 490 nm.

Serum samples

Several collections of specimens were used to characterize the various fragments and to assess the NC Mosaic EIA: 1.) 128 anti-HCV positive specimens obtained from paid plasma donors (Boston Biomedica Inc., West Bridgewater, MA), 2.) a collection of normal blood donors negative for anti-HCV activity repositied at CDC, 3.) 21 anti-HCV positive and genotyped specimens (Boehringer Mannheim, Mannheim, Germany), and 4.) 4 anti-HCV positive seroconversion panels (Serologicals Inc., Clarkston, GA).

NC Mosaic EIA Results

A frequency distribution of 200 anti-HCV negative specimens were tested by EIA to statistically derive a cutoff value (Figure 12). This value was set at an OD value greater than the mean OD plus 3.5 standard deviations of the mean or 0.145. When applying this cutoff value one of the anti-HCV negative specimens gave an OD value equal to 0.145, which was interpreted as negative, giving an overall specificity of 100%.

Two serially diluted specimens, BBI 304 and BBI 325 were tested by the NC Mosaic EIA and by MATRIX to determine relative sensitivities. The results were expressed as sample to cutoff values (S/CO) so that each test may be directly compared (Figures 13 and 14, respectively). A S/CO value greater than 1 is considered positive. Specimen BBI 325 reached an endpoint by MATRIX at a dilution of 1:256,000. NC Mosaic EIA gave a S/CO value of 1.8 at that dilution; however, an examination of cutoff values at a 1:64,000 dilution and at a 1:128,000 dilution suggests that the S/CO value for the EIA may not be accurate and that the true endpoint by NC Mosaic EIA may be at a dilution of 1:32,000

or 4-fold less sensitive than MATRIX. Conversely, Specimen BBI 304 gave an endpoint titer of 1:128,000 by MATRIX, while the NC Mosaic EIA was still positive at a dilution of 1:256,000 suggesting that the EIA was 2-fold more sensitive than MATRIX. It is not unusual for several samples to give different endpoint titers since the immunologic targets are very different. The endpoint titers obtained by these two assays on the same sera most probably is a reflection of the relative titers of antibodies to different antigenic epitopes as they are presented within each test format.

To measure clinical sensitivity several seroconversion panels (No. 4812, 4813, and 4814) were tested by the NC Mosaic EIA and by MATRIX (Figures 15, 16, and 17, respectively). A cutoff value of 2.5 times background was used for the NC Mosaic EIA, while a S/CO value greater than 1.0 was used for MATRIX. All three seroconversion panels detected anti-NC activity at approximately the same number of days after transfusion.

Another manner to measure clinical sensitivity is to test a panel of anti-HCV positive sera for anti-NC activity by NC Mosaic EIA and by MATRIX. A panel of 128 specimens obtained from professional plasma donors tested positive by a commercially available EIA screening assay. Among the 128 initially reactive specimens, 109 were confirmed as positive by MATRIX, while 12 tested as indeterminate and 7 as negative. Among the 109 confirmed anti-HCV positive specimens, 101 (92.6%) demonstrated anti-NC activity by MATRIX and 99 (90.8%) by NC Mosaic EIA suggesting a slightly higher sensitivity for MATRIX. Among the 12 indeterminate specimens, 6 demonstrated anti-NC activity by MATRIX, and 3 by NC Mosaic EIA suggesting a higher specificity for the NC Mosaic EIA. None of the 7 anti-HCV negatives were positive for anti-NC activity by either test. (Data not shown).

In another study, among 78 initially reactive specimens 66 were confirmed as anti-HCV positive by MATRIX, one specimen tested indeterminate, while 3 tested as negative. The NC Mosaic EIA gave concordant results with MATRIX for anti-NC activity for the 66 positive samples and for the one negative

specimen. The indeterminate specimen tested negative for anti-NC activity by NC Mosaic EIA suggesting a higher specificity for this specimen. The remaining 8 specimens were known to have nonspecific reactivity to the NS4 antigen, but tested negative by both assays for anti-NC activity. (Data not shown).

Finally, 23 anti-HCV sera representing genotypes 1 - 5 were tested for anti-NC activity by NC Mosaic EIA and by MATRIX. The results indicating a 100% concordance between the two assays (data not shown) indicating that the mosaic NC protein, although composed of sequences from genotypes 1 - 3, contains crossreacting epitopes that react with anti-NC positive sera obtained from individuals infected with 5 different genotypes. Collectively, these results suggest that the NC mosaic protein when used as the immunologic target in an EIA format is at least as sensitive and possibly more specific than MATRIX for the detection of anti-NC activity.

Example 3

Design and production of an artificial NS4 mosaic protein.

Design of an artificial NS4 mosaic protein

To construct an artificial NS4 antigen containing antigenic epitopes from several HCV genotypes, all sequences from the 5-1-1 region as well as a strongly immunoreactive region located at the C-terminal of NS4 were searched in GeneBank. Representative regions from different genotypes were selected based upon significant sequence divergence from each other and are shown in Figure 18.

Gene Assembly

The amino acid sequence as shown in Figure 18 was back translated into its nucleotide sequence, and synthetic oligonucleotides were used to construct nine monomers, the first eight consisting of two antigenic domains and the ninth consisting of one antigenic domain, and were consecutively assembled using Restriction Endonuclease Assisted Ligation (REAL). The synthetic oligonucleotides used were as follows:

A1: 5'- CCC CGA ATT CAA GCC GCC CAC ATA CCA TAC
CTA GAA CAA GGA ATG CAT CTC GCA GAA CAA TTC
AAA CAA AAG GCA CTT CGT C (SEQ ID NO:34);

A2: 5'- CCC CGG ATC CTA ACT AGC CTC TTC CAT CTC
ATC AAA CTC CTG ATA CAA AAC CTC CCT ATC CGG
GAT AAC AGC CGG ACG AAG TGC (SEQ ID NO:35);

B1: 5'- CCC CGA ATT CAA GCT AGT CAC TTA CCG TAT
ATC GAG CAG GGA ATG CAG TTA GCT GAA CAG TTT
AAA CAG AAG GCT CTG GCT TTT G (SEQ ID NO:36);

B2: 5'- CCC CGG ATC CTA AGG CCG AGC GTC AGA CTC
AGG AAC ATA ATG AGT AGG AGA AAC ATG ATT ACC
CCG AGA AGC AAA AGC CAG (SEQ ID NO:37);

C1: 5'- CCC CGA ATT CAA CGG CCT GCG ATA ATA CCG
GAT AGG GAG GTT CTT CAT AGG GAG TTT GAC GAG
ATG GAG GAG GCT TTT GCG (SEQ ID NO:38);

C2: 5'- CCC CGG ATC CTA CTG CGA AGC ATC AGA CTC
AGG AAC ATA ATG AGC CGG ACT AAC ATG ATT CCC
ACG AGA CGC AAA AGC C (SEQ ID NO:39);

D1: 5'- CCC CGA ATT CAA TCG CAG GCG GCG CCT TAT
ATT GAG CAG GCT CAG GTT ATT GCT CAT CAG TTT
AAG GAG AAG GTT CTT GCT TT (SEQ ID NO:40);

D2: 5'- CCC CGG ATC CTA CGG CTT CGC GTC CGA CTC
AGG AAC ATA ATG AGT CGG AGA ATC ATG ATT ACC
ACG AGA AGC AAA AGC AAG AA (SEQ ID NO:41);

E1: 5'- CCC CGA ATT CAA AAG CCG GCG ATA ATC CCT
GAC CGT GAG GTT CTG TAT CGT GAG TTT GAT GAG
ATG GAG GAG TCA CAG C (SEQ ID NO:42);

E2: 5'- CCC CGG ATC CTA AAA CGC CAG AGC CTT CTG
CTT AAA CTG CTC AGC AAG CAT CAT ACC CTG CTC AAT
GTA CGG AAG ATG CTG TGA CTC (SEQ ID NO:43);

F1: 5'- CCC CGA ATT CAA GCG TTT GCT TCT CGT GGT
AAT CAT GTT GCT CCG ACT CAT TAT GTT ACG GAG
TCA GAT GCT AAG C (SEQ ID NO:44);

F2: 5'- CCC CGG ATC CTA GAA AGC CTC CTC CAT CTC
ATC ATA CTG CTG ATA AAG AAC CTC CTT ATC CGG
AAC CAG AGC CGG CTT AGC ATC (SEQ ID NO:45);

G1: 5'- CCC CGA ATT CAA GCT TTC GCT TCT CGT GGT
AAT CAT GTT GCT CCT ACG CAT TAT GTT GTT GAG TCA
GAT GCT TCT GCT TC (SEQ ID NO:46);

G2: 5'- CCC CGG ATC CTA GAA AGC CAG AAC CTT CTC
CTT AAA CTG ACC AGC AAT AGC ACG CGT CTC GTC
CAT ATA CGG CAG AGA AGC AGA AG (SEQ ID NO:47);

H1: 5'- CCC CGA ATT CAA GCT TTC GCT AGT CGT GGG
AAT CAT GTG TCG CCG CGT CAT TAT GTG CCT GAG TCT
GAG CCT CAG GTT GT (SEQ ID NO:48);

H2: 5'- CCC CGG ATC CTA AGA AGC CTC CTC CAT CTC
ATC AAA AGC CTC ATA CAG TAT CTC CTT ATC CGG CGT
AAC AAC AAC CTG AG (SEQ ID NO:49);

I1: 5'- CCC CGA ATT CAA GCT TCT AAG GCC GCG CTG
ATT GAG GAG GGT CAG CGT ATG G (SEQ ID NO:50);

I2: 5'- CCC CGG ATC CTA CTG GAT CTT AGA CTT CAG
CAT CTC AGC CAT ACG CTG (SEQ ID NO:51);

Gene expression and Protein Purification

To express the synthetic genes, *E. coli* JM109 competent cells (Promega, Madison, WI) were transformed with

plasmids containing 9 monomers, 4 dimers, 2 tetramers, and a full size gene using the REAL method described in Example 1. Cells were grown in Luria broth (LB) with 50 mg/ml ampicillin overnight at 37°C. The cultures were then diluted 1:10 in fresh LB with 50 mg/ml ampicillin and grown 3 to 4 hours until the optical density at 600 nm reached 0.5 - 1.0. The gene was expressed by activating the tac promoter by the addition of isopropyl-b-D-thiogalactoside (IPTG, Sigma Chemical Co., St. Louis, MO) at a final concentration of 1 mM. Cells were harvested 30 minutes after induction at 37°C. Cell lysates were prepared and the soluble fraction of the lysates was obtained by centrifugation at 12,000xg for 20 minutes. The glutathione S-transferase (GST) - mosaic proteins were purified by affinity chromatography using glutathione sepharose 4B column (Pharmacia Biotech, Piscataway, NJ).

Immunoblot Assay

Aliquots of each lysate or aliquots of the purified GST-mosaic proteins were subjected to electrophoresis on 12% polyacrylamide gels containing SDS (SDS-PAGE) followed by blotting onto a nitrocellulose membrane. Following protein transfer, the nitrocellulose membranes were incubated with blocking solution (0.1 M phosphate-buffered saline containing 1% bovine serum albumin, 0.5% Tween 20, and 10% normal goat serum) overnight at 4°C, and then incubated with human HCV positive sera diluted 1:100 or 1:200 in blocking solution for 1 hour at room temperature. For immunodetection, the membranes were washed three times with blocking solution, followed by the addition of affinity-purified goat anti-human immunoglobulin G (IgG) conjugated to horseradish peroxidase (Bio-Rad, Richmond, CA) diluted 1:4000 or 1:6000 in blocking solution, and incubated 1 hour at room temperature. After washing the membranes with blocking solution three times, diaminobenzidine (Sigma Chemical Co., St. Louis, MO) and hydrogen peroxidase were added to detect the presence of the horseradish peroxidase (HRP) reporter molecule.

Enzyme Immunoassay (EIA)

One hundred microliters of the purified full length fusion NS4 mosaic protein (GST-W3) was adsorbed to microtiter wells (Immuno II; Dynatech Laboratories, Inc., Chantilly, VA) at a concentration of 100 ng/ml in 0.1 M phosphate-buffered saline, pH 7.5, overnight at room temperature. The microtiter wells were then incubated with human anti-HCV negative or positive sera diluted 1:500 in blocking solution (as described above for the immunoblot assay) for 1 hour at 37°C. After washing the microtiter wells, goat anti-human immunoglobulin G (IgG) conjugated to HRP diluted 1:4000 was added and incubated for 1 hour at 37°C. After washing the microtiter wells 5 times, substrate and chromophore was added (Abbott Diagnostics Division, North Chicago, IL) and incubated in the dark for 30 minutes. The reaction was stopped with acid and the optical density was measured at 493 nm.

Human sera

Anti-HCV positive sera were obtained from Boehringer Mannheim Inc. (Penzberg, Germany) and from Boston Biomedical Inc. (West Bridgewater, MA). Anti-HCV negative sera were obtained from a collection of normal human blood donors deposited at the Centers for Disease Control and Prevention (CDC, Atlanta, GA). All sera were confirmed as anti-HCV positive or negative by EIA and initially reactive specimens were confirmed and further characterized by the supplemental test MATRIX (Abbott Laboratories, Abbott Park, IL).

Gene Assembly

The full length artificial NS4 mosaic gene was constructed in sequential steps from synthetic oligonucleotides by REAL. As shown in Figure 19, each pair of oligonucleotides were converted into 9 monomers (A, B, C, D, E, F, G, H, and I), which were then consecutively assembled into 4 dimers (AB, CD, EF, and GH). Consecutive dimers were then assembled into 2 tetramers (ABCD and EFGH). The final gene was assembled from the 2 tetramers and the remaining monomer (I).

Gene Expression and Immunoblot Assay

SDS-PAGE analysis demonstrated that each fragment (9 monomers, 4 dimers, 2 tetramers, and the full length gene) was expressed at high levels following induction for 5 hours at 37°C with 1 mM IPTG. Each of the expressed fragments and the expressed full length gene were purified by ligand affinity chromatography. All of the purified proteins were shown to be highly purified by SDS-PAGE, although many of the purified proteins displayed an artifactual doublet. In addition, each of the purified proteins were analyzed by immunoblot (Figure 20) to ascertain individual immunoreactivity to human anti-HCV positive sera. The immunoblot showed that most of the purified proteins were strongly immunoreactive with a single anti-HCV positive specimen diluted 1:200. Three monomers (A, H, and I), however, were not immunoreactive using this specimen diluted 1:200. Monomers A and H were immunoreactive using pooled sera diluted 1:10 indicating that these monomers were immunoreactive. Monomer I demonstrated weak immunoreactivity by EIA.

NS4 Mosaic EIA Frequency Distribution

A statistically valid cutoff value was determined by screening 160 anti-HCV negative sera and 166 anti-HCV positive sera (anti-NS4 positive by MATRIX) by EIA. The results showed that approximately 90% of anti-HCV negative sera gave OD values less than 0.09, while approximately 80% of anti-HCV positive sera gave OD values greater than 2.1. The mean OD value for the anti-HCV negative specimens was 0.0518 ± 0.0273 standard deviations (SD). The cutoff value was established as the mean of OD values for anti-HCV negative sera plus 3.5 times the SD of the mean. This cutoff value unambiguously separated the negative sera from the positive sera (Figure 21), although one negative specimen gave an OD value slightly above this cutoff. Using this cutoff value, all of the anti-HCV positive specimens tested positive by the NS4 Mosaic EIA. A two by two analysis of the data revealed a sensitivity of 100% and a specificity of 99.4% using this derived

cutoff value. By raising the cutoff to the mean + 4.3 times the SD, the specificity compared to MATRIX was 100%.

NS4 Mosaic EIA Compared to MATRIX on Serially Diluted Anti-HCV Positive Sera

To examine the antigenic reactivity of the NS4 mosaic protein in detecting anti-NS4 activity, two serially diluted anti-NS4 positive sera were tested by the NS4 Mosaic EIA and by MATRIX. The results showed that anti-NS4 antibody can be detected by the NS4 Mosaic EIA at a dilution of 1:128,000 times, while MATRIX was positive for anti-NS4 activity at a dilution of approximately 1:4000. MATRIX utilizes two different NS4 proteins expressed in *E. coli* and in yeast. This comparison indicated that the antigenic reactivity to the NS4 mosaic protein was 32 times more sensitive than MATRIX for specimen no. 1 (Figure 22A) and 18 to 25 times more sensitive for specimen no. 2 (Figure 22B).

NS4 Mosaic EIA Compared to MATRIX for the Detection of Anti-HCV

Among 182 anti-HCV positive sera, 97.8% tested positive for anti-NS4 activity by the NS4 Mosaic EIA compared to 86.8% by MATRIX. These results strongly suggest that the mosaic protein is a more sensitive immunologic target than either of the NS4 antigens used by MATRIX. Antibody activity to the NS3 and nucleocapsid (NC) antigens by MATRIX were also compared to the mosaic protein for anti-NS4 activity. This analysis showed that 98.4% of the 182 sera tested positive for anti-NS3 and 94.5% for anti-NC indicating that the NS4 Mosaic EIA is more sensitive than MATRIX for anti-NC activity, and almost as sensitive as MATRIX for anti-NS3 activity (Figure 23).

NS4 Mosaic EIA Compared to MATRIX for Seroconversion Panels

Ten seroconversion panels (BioClinical Partners, Inc.; Serologicals, Chamblee, GA) were tested by the NS4 Mosaic EIA and by MATRIX to determine the temporal appearance of anti-NS4 activity in recently infected individuals. The results showed

that the NS4 Mosaic EIA detected anti-NS4 activity approximately 15 (Figure 24) to 25 days (Figure 24) earlier than MATRIX when a cutoff value of at least 2.5 times background was used. In some cases, the NS4 Mosaic EIA and MATRIX gave similar results; however, MATRIX results never demonstrated earlier detection of anti-NS4 activity than NS4 Mosaic EIA results (data not shown). These results indicate that the NS4 mosaic protein, when used as the immunologic target in an EIA, was at least as sensitive as MATRIX for the early detection of anti-NS4 activity, and probably more sensitive if more frequent bleed dates were available for each of the ten seroconversion panels.

NS4 Mosaic EIA Reactivity to Different HCV Genotypes

Since the NS4 mosaic protein is composed of antigenic regions derived from several HCV subtypes and genotypes, it should detect anti-NS4 activity in the sera from patients infected with different genotypes. Genotypes 1-5 were tested for immunoreactivity by the NS4 Mosaic EIA. The results indicated that the only specimens which did not react to the mosaic protein were those that tested negative for anti-NS4 activity by MATRIX. These data indicate that the mosaic protein detected anti-NS4 activity in each of the genotypes tested and was 100% concordant with MATRIX (Figure 25).

SEQUENCE LISTING

(1) GENERAL INFORMATION:

- (i) APPLICANT: KHUDYAKOV, YURI E.
FIELDS, HOWARD A.
- (ii) TITLE OF INVENTION: MOSAIC PROTEIN AND RESTRICTION
ENDONUCLEASE ASSISTED LIGATION METHOD FOR MAKING
THE SAME
- (iii) NUMBER OF SEQUENCES: 55
- (iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: JONES & ASKEW, LLP
(B) STREET: 191 Peachtree Street, N.W., 37th Floor
(C) CITY: Atlanta
(D) STATE: GA
(E) COUNTRY: USA
(F) ZIP: 30303-1769
- (v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30
- (vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER: US
(B) FILING DATE:
(C) CLASSIFICATION:
- (viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: WARREN, WILLIAM L.
(B) REGISTRATION NUMBER: 36,714
(C) REFERENCE/DOCKET NUMBER: 03063-0380
- (ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: 404-818-3700
(B) TELEFAX: 404-818-3799

(2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 55 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cdNA
- (iii) HYPOTHETICAL: NO

- (iv) ANTI-SENSE: NO
- (v) FRAGMENT TYPE: internal
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Hepatitis virus

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

CCCCGAATTC AACCGAAACC GCAACGTAAA ACCAAACGTA ACACCATTTCG TCGTC
55

(2) INFORMATION FOR SEQ ID NO:2:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 69 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO
- (v) FRAGMENT TYPE: internal
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Hepatitis virus

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

CCCCGGATCC TATTTCGGAC CAACGATCTG ACCACCACCC GGGAATTTAA
CGTCCTGCGG 60

ACGACGAAT
69

(2) INFORMATION FOR SEQ ID NO:3:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 54 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO

- (iv) ANTI-SENSE: NO
- (v) FRAGMENT TYPE: internal
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Hepatitis virus

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

CCCCGAATTC AACCGAAACC GCAACGTCAG ACCAAACGTA ACACCAACCG TCGT
54

(2) INFORMATION FOR SEQ ID NO:4:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 70 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO
- (v) FRAGMENT TYPE: internal
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Hepatitis virus

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

CCCCGATCC TATTTCCGGAC CAACGATCTG ACCACCACCC GCGAATTTAA
CGTCCTGCGG 60

ACGACGGTTG
70

(2) INFORMATION FOR SEQ ID NO:5:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 51 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO

- (iv) ANTI-SENSE: NO
- (v) FRAGMENT TYPE: internal
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Hepatitis virus

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

CCCCGAATTC AACCGAAACC GCAACGTAAA ACCAAACGTA ACACCTACCG T
51

(2) INFORMATION FOR SEQ ID NO:6:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 73 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO
- (v) FRAGMENT TYPE: internal
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Hepatitis virus

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

CCCCGGATCC TATTTCCGGAC CAACGATCTG ACCACCACCC GCGAATTAA
CGTCCTGCGG 60

ACGACGGTAG GTG
73

(2) INFORMATION FOR SEQ ID NO:7:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 55 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO

- (iv) ANTI-SENSE: NO
- (v) FRAGMENT TYPE: internal
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Hepatitis virus

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

CCCCGAATTC AACCGAAACC GCAACGTAAA CCGAACCGTA ACACCAACCG TCGTC
55

(2) INFORMATION FOR SEQ ID NO:8:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 69 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO
- (v) FRAGMENT TYPE: internal
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Hepatitis virus

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

CCCCGGATCC TATTTCGGAC CAACGATCTG ACCACCACC GCGAATTTAA
CGTCCTGCGG 60

ACGACGGTT
69

(2) INFORMATION FOR SEQ ID NO:9:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 63 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO

- (iv) ANTI-SENSE: NO
- (v) FRAGMENT TYPE: internal
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Hepatitis virus

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

CCCCGAATTC AACCGAAACC GCAACGTCAG CCGAAACGTA ACACCCCGCG
TCGTCCGCAG 60

GAC
63

(2) INFORMATION FOR SEQ ID NO:10:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 60 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO
- (v) FRAGMENT TYPE: internal
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Hepatitis virus

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

CCCCGGATCC TATTTCGGAC CAACGATCTG ACCACCACCC GGGAATTAA
CGTCCTGCGG 60

(2) INFORMATION FOR SEQ ID NO:11:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 55 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO

- (iv) ANTI-SENSE: NO
- (v) FRAGMENT TYPE: internal
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Hepatitis virus

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

CCCCGAATTC AACCGAAACC GCAACGTAAA ACCAAACGTA ACGCTCACCG TCGTC
55

(2) INFORMATION FOR SEQ ID NO:12:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 69 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO
- (v) FRAGMENT TYPE: internal
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Hepatitis virus

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

CCCCGGATCC TATTTCGGAC CAACGATCTG ACCACCACCC GGGAATTTAA
CGTCCTGCGG 60

ACGACGGTG
69

(2) INFORMATION FOR SEQ ID NO:13:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 55 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO

- (iv) ANTI-SENSE: NO
- (v) FRAGMENT TYPE: internal
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Hepatitis virus

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

CCCCGAATTC AACCGAAACC GCAACGTAA AACCAGCGTA ACACCAACCG TCGTC
55

(2) INFORMATION FOR SEQ ID NO:14:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 69 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO
- (v) FRAGMENT TYPE: internal
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Hepatitis virus

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

CCCCGGATCC TATTTCGGAC CAACGATCTG ACCACCACCC GCGAATTTAA
CGTCCTGCGG 60

ACGACGGTT
69

(2) INFORMATION FOR SEQ ID NO:15:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 55 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO

005270-5116460

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

(2) INFORMATION FOR SEQ ID NO:16:

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

```
(v) FRAGMENT TYPE: internal
```

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Hepatitis virus

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

CCCCGGATCC TATTTCTGGAA CGTAGATAAC ACCACCACCC GGGAATTTAA
CGTCCTGCGG 60

ACGACGAAT
69

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 56 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

- (iv) ANTI-SENSE: NO
- (v) FRAGMENT TYPE: internal
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Hepatitis virus

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

CCCCGAATTC AACCGAAACC GCAACGTAAA ACCGAACGTA ACACCAACCG TCGTCC
56

(2) INFORMATION FOR SEQ ID NO:18:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 65 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO
- (v) FRAGMENT TYPE: internal
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Hepatitis virus

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

CCCCGATCC TATTTCCGAC CAACGATCTG ACCACCACCA GAGAAACGAA
CGTCCGGACG 60

ACGGT
65

(2) INFORMATION FOR SEQ ID NO:19:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 54 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

CCCCGAATTC AACCGAAACC GAAACGTCAG ACCAAACGTA ACACCCTGCG TCGT
54

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 73 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(v) FRAGMENT TYPE: internal

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Hepatitis virus

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

CCCCGGATCC TATTTCGGAC CAACGATCTG ACCACCAGCC GGGAATTTAA
CGTTTTTCGG 60

ACGACGACGC AGG
73

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 55 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

(2) INFORMATION FOR SEQ ID NO:22:

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(v) FRAGMENT TYPE: internal

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Hepatitis virus

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

CCCCGGATCC TATTTCGGAC CAACGATCTG ACCACCACCC GGGAATTTAA
CGTCCTGCGG 60

ACGACGGTG
69

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 28 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: not relevant
- (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Hepatitis virus

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

Pro	Lys	Pro	Gln	Arg	Lys	Thr	Lys	Arg	Asn	Thr	Ile	Arg	Arg
1					5				10				15
Asp	Val	Lys	Phe	Pro	Gly	Gly	Gly	Gln	Ile	Val	Gly		
			20					25					

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 28 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: not relevant

(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(v) FRAGMENT TYPE: internal

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Hepatitis virus

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

Pro	Lys	Pro	Gln	Arg	Gln	Thr	Lys	Arg	Asn	Thr	Asn	Arg	Arg
1					5				10				15
Asp	Val	Lys	Phe	Pro	Gly	Gly	Gly	Gln	Ile	Val	Gly		
			20					25					

(2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 28 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: not relevant

(D) TOPOLOGY: not relevant

- (ii) MOLECULE TYPE: protein
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO
- (v) FRAGMENT TYPE: internal
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Hepatitis virus

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

Pro	Lys	Pro	Gln	Arg	Lys	Thr	Lys	Arg	Asn	Thr	Tyr	Arg	Arg
Pro	Gln												
1				5					10				15
Asp	Val	Lys	Phe	Pro	Gly	Gly	Gly	Gln	Ile	Val	Gly		
			20					25					

(2) INFORMATION FOR SEQ ID NO:26:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 28 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: not relevant
 - (D) TOPOLOGY: not relevant
- (ii) MOLECULE TYPE: protein
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO
- (v) FRAGMENT TYPE: internal
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Hepatitis virus

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

Pro	Lys	Pro	Gln	Arg	Lys	Pro	Asn	Arg	Asn	Thr	Asn	Arg	Arg
Pro	Gln												
1				5					10				15
Asp	Val	Lys	Phe	Pro	Gly	Gly	Gly	Gln	Ile	Val	Gly		
			20					25					

(2) INFORMATION FOR SEQ ID NO:27:

- (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 28 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: not relevant
- (D) TOPOLOGY: not relevant

- (ii) MOLECULE TYPE: protein
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO
- (v) FRAGMENT TYPE: internal
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Hepatitis virus

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

Pro Lys Pro Gln Arg Gln Pro Lys Arg Asn Thr Pro Arg Arg
 Pro Gln 1 5 10 15
 Asp Val Lys Phe Pro Gly Gly Gly Gln Ile Val Gly
 20 25

- (2) INFORMATION FOR SEQ ID NO:28:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 28 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: not relevant
 - (D) TOPOLOGY: not relevant

- (ii) MOLECULE TYPE: protein
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO
- (v) FRAGMENT TYPE: internal
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Hepatitis virus

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

Pro Lys Pro Gln Arg Lys Thr Lys Arg Asn Ala His Arg Arg
 Pro Gln 1 5 10 15
 Asp Val Lys Phe Pro Gly Gly Gly Gln Ile Val Gly

0041116 075500

20

25

(2) INFORMATION FOR SEQ ID NO:29:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 28 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: not relevant
- (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(v) FRAGMENT TYPE: internal

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Hepatitis virus

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

Pro	Lys	Pro	Gln	Lys	Arg	Asn	Gln	Arg	Asn	Thr	Asn	Arg	Arg
Pro	Gln												
1				5				10					15
Asp	Val	Lys	Phe	Pro	Gly	Gly	Gly	Gln	Ile	Val	Gly		
			20					25					

(2) INFORMATION FOR SEQ ID NO:30:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 28 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: not relevant
- (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(v) FRAGMENT TYPE: internal

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Hepatitis virus

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

Pro Lys Pro Gln Arg Lys Thr Lys Arg Asn Thr Ile Arg Arg
 Pro Gln 1 5 10 15
 Asp Val Lys Phe Pro Gly Gly Gly Val Ile Tyr Val
 20 25

(2) INFORMATION FOR SEQ ID NO:31:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 28 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: not relevant
 - (D) TOPOLOGY: not relevant

- (ii) MOLECULE TYPE: protein

- (iii) HYPOTHETICAL: NO

- (iv) ANTI-SENSE: NO

- (v) FRAGMENT TYPE: internal

- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Hepatitis virus

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

Pro Lys Pro Gln Arg Lys Thr Glu Arg Asn Thr Asn Arg Arg
 Pro Gln 1 5 10 15
 Asp Val Arg Phe Ser Gly Gly Gly Gln Ile Val Gly
 20 25

(2) INFORMATION FOR SEQ ID NO:32:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 28 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: not relevant
 - (D) TOPOLOGY: not relevant

- (ii) MOLECULE TYPE: protein

- (iii) HYPOTHETICAL: NO

- (iv) ANTI-SENSE: NO

- (v) FRAGMENT TYPE: internal

- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Hepatitis virus

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

Pro Lys Pro Lys Arg Gln Thr Lys Arg Asn Thr Leu Arg Arg
 Pro Lys 1 5 10 15
 Asn Val Lys Phe Pro Ala Gly Gly Gln Ile Val Gly
 20 25

(2) INFORMATION FOR SEQ ID NO:33:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 28 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: not relevant
 - (D) TOPOLOGY: not relevant

- (ii) MOLECULE TYPE: protein

- (iii) HYPOTHETICAL: NO

- (iv) ANTI-SENSE: NO

- (v) FRAGMENT TYPE: internal

- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Hepatitis virus

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

Pro Lys Pro Gln Arg Lys Thr Lys Arg Lys Ala His Arg Arg
 Pro Gln 1 5 10 15
 Asp Val Lys Phe Pro Gly Gly Gly Gln Ile Val Gly
 20 25

(2) INFORMATION FOR SEQ ID NO:34:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 82 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: cDNA

- (iii) HYPOTHETICAL: NO

- (iv) ANTI-SENSE: NO

(v) FRAGMENT TYPE: internal

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Hepatitis virus

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

CCCCGAATTC AAGCCGCCCA CATACCATAC CTAGAACAAG GAATGCATCT
CGCAGAACAA 60

TTCAAACAAA AGGCACTTCG TC
82

(2) INFORMATION FOR SEQ ID NO:35:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 84 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(v) FRAGMENT TYPE: internal

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Hepatitis virus

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

CCCCGGATCC TAACTAGCCT CTTCATCTC ATCAAATCCTC TGATACAAAA
CCTCCCTATC 60

CGGGATAACA GCCGGACGAA GTGC
84

(2) INFORMATION FOR SEQ ID NO:36:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 85 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

005270-001100

- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO
- (v) FRAGMENT TYPE: internal
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Hepatitis virus

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

CCCCGAATTC AAGCTAGTCA CTTACCGTAT ATCGAGCAGG GAATGCAGTT
AGCTGAACAG 60

TTTAAACAGA AGGCTCTGGC TTTTG
85

(2) INFORMATION FOR SEQ ID NO:37:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 81 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO
- (v) FRAGMENT TYPE: internal
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Hepatitis virus

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

CCCCGGATCC TAAGCCCGAG CGTCAGACTC AGGAACATAA TGAGTAGGAG
AAACATGATT 60

ACCCCGAGAA GCAAAAGCCA G
81

(2) INFORMATION FOR SEQ ID NO:38:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 81 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO
- (v) FRAGMENT TYPE: internal
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Hepatitis virus

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

CCCCGAATTC AACGGCCTGC GATAATACCG GATAGGGAGG TTCTTCATAG
GGAGTTTGAC 60

GAGATGGAGG AGGCTTTTGC G
81

(2) INFORMATION FOR SEQ ID NO:39:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 79 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO
- (v) FRAGMENT TYPE: internal
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Hepatitis virus

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

CCCCGGATCC TACTGCGAAG CATCAGACTC AGGAACATAA TGAGCCGGAC
TAACATGATT 60

CCCACGAGAC GCAAAGGCC
79

(2) INFORMATION FOR SEQ ID NO:40:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 83 base pairs

004116-02500

- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO
- (v) FRAGMENT TYPE: internal
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Hepatitis virus

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

CCCCGAATTC AATCGCAGGC GCGCCTTAT ATTGAGCAGG CTCAGGTTAT
TGCTCATCAG 60

TTTAAGGAGA AGGTCTTGC TTT
83

(2) INFORMATION FOR SEQ ID NO:41:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 83 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO
- (v) FRAGMENT TYPE: internal
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Hepatitis virus

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:

CCCCGGATCC TACGGCTTCG CGTCCGACTC AGGAACATAA TGAGTCGGAG
AATCATGATT 60

ACCACGAGAA GCAAAAGCAA GAA
83

(2) INFORMATION FOR SEQ ID NO:42:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 79 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(v) FRAGMENT TYPE: internal

- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Hepatitis virus

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:

CCCCGAATTC AAAAGCCGGC GATAATCCCT GACCGTGAGG TTCTGTATCG
TGAGTTTGAT 60

GAGATGGAGG AGTCACAGC
79

(2) INFORMATION FOR SEQ ID NO:43:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 87 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(v) FRAGMENT TYPE: internal

- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Hepatitis virus

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:

CCCCGGATCC TAAACGCCA GAGCCTTCTG CTTAAACTGC TCAGCAAGCA
TCATACCCTG 60

00000-00000

CTCAATGTAC GGAAGATGCT GTGACTC
87

(2) INFORMATION FOR SEQ ID NO:44:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 76 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(v) FRAGMENT TYPE: internal

- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Hepatitis virus

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:44:

CCCCGAATTC AAGCGTTTGC TTCTCGTGGT AATCATGTTG CTCCGACTCA
TTATGTTACG 60

GAGTCAGATG CTAAGC
76

(2) INFORMATION FOR SEQ ID NO:45:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 84 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(v) FRAGMENT TYPE: internal

- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Hepatitis virus

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:45:

CGGAACCAGA GCCGGCTTAG CATC
84

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 80 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(v) FRAGMENT TYPE: internal

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Hepatitis virus

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:46:

CCCCGAATTC AAGCTTTCGC TTCTCGTGGT AATCATGTTG CTCCTACGCA
TTATGTTGTT 60

GAGTCAGATG CTTCTGCTTC
80

(2) INFORMATION FOR SEQ ID NO:47:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 86 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(v) FRAGMENT TYPE: internal

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Hepatitis virus

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:47:

CCCCGGATCC TAGAAAGCCA GAACCTTCTC CTTAAACTGA CCAGCAATAG
CACGCGTCTC 60

GTCCATATAC GGCAGAGAAG CAGAAG
86

(2) INFORMATION FOR SEQ ID NO:48:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 80 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO
- (v) FRAGMENT TYPE: internal
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Hepatitis virus

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:48:

CCCCGAATTC AAGCTTTCGC TAGTCGTGGG AATCATGTGT CGCCGCGTCA
TTATGTGCCT 60

GAGTCTGAGC CTCAGGTTGT
80

(2) INFORMATION FOR SEQ ID NO:49:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 80 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO
- (v) FRAGMENT TYPE: internal
- (vi) ORIGINAL SOURCE:

00220-944660

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:49:

CGGCGTAACA ACAACCTGAG
80

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 52 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(iii) HYPOTHETICAL: NO

(v) FRAGMENT TYPE: internal

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Hepatitis virus

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:50:

CCCCGAATTC AAGCTTCTAA GGCCGCGCTG ATTGAGGAGG GTCAGCGTAT GG
52

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 48 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(iii) HYPOTHETICAL: NO

(v) FRAGMENT TYPE: internal

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:51:

(2) INFORMATION FOR SEQ ID NO:52:

(ii) MOLECULE TYPE: protein

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:52:

Glu Ser Asp Ala Arg Pro Ala Ile Ile Pro Asp Arg Glu Val
Leu His

	85	90	95
Arg Glu Phe Asp Glu Met Glu Glu Ala Phe Ala Ser Arg Gly			
Asn His	100	105	110
Val Ser Pro Ala His Tyr Val Pro Glu Ser Asp Ala Ser Gln			
Ala Ala	115	120	125
Pro Tyr Ile Glu Gln Ala Gln Val Ile Ala His Gln Phe Lys			
Glu Lys	130	135	140
Val Leu Ala Phe Ala Ser Arg Gly Asn His Asp Ser Pro Thr			
His Tyr	145	150	155
160			
Val Pro Glu Ser Asp Ala Lys Pro Ala Ile Ile Pro Asp Arg			
Glu Val	165	170	
175			
Leu Tyr Arg Glu Phe Asp Glu Met Glu Glu Ser Gln His Leu			
Pro Tyr	180	185	190
Ile Glu Gln Gly Met Met Leu Ala Glu Gln Phe Lys Gln Lys			
Ala Leu	195	200	205
Ala Phe Ala Ser Arg Gly Asn His Val Ala Pro Thr His Tyr			
Val Thr	210	215	220
Glu Ser Asp Ala Lys Pro Ala Leu Val Pro Asp Lys Glu Val			
Leu Tyr	225	230	235
240			
Gln Gln Tyr Asp Glu Met Glu Glu Ala Phe Ala Ser Arg Gly			
Asn His	245	250	
255			
Val Ala Pro Thr His Tyr Val Val Glu Ser Asp Ala Ser Ala			
Ser Leu	260	265	270
Pro Tyr Met Asp Glu Thr Arg Ala Ile Ala Gly Gln Phe Lys			
Glu Lys	275	280	285

Val Leu Ala Phe Ala Ser Arg Gly Asn His Val Ser Pro Arg
 His Tyr 290 295 300

Val Pro Glu Ser Glu Pro Gln Val Val Val Thr Pro Asp Lys
 Glu Ile 305 310 315
 320

Leu Tyr Glu Ala Phe Asp Glu Met Glu Glu Ala Ser Lys Ala
 Ala Leu 325 330
 335

Ile Glu Glu Gly Gln Arg Met Ala Glu Met Leu Lys Ser Lys
 Ile Gln 340 345 350

(2) INFORMATION FOR SEQ ID NO:53:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 57 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO
- (v) FRAGMENT TYPE: internal
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Hepatitis virus

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:53:

CTGGTTCCGC GTGGATCCCC AGGAATTCCT GGGTCGACTC GAGCGGCCGC ATCGTGA
 57

(2) INFORMATION FOR SEQ ID NO:54:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 27 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA

005210-54116460

- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO
- (v) FRAGMENT TYPE: internal
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Hepatitis virus

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:54:

TCGCAGCGAA TTCTCGAGGA TCCATCC
27

(2) INFORMATION FOR SEQ ID NO:55:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 56 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO
- (v) FRAGMENT TYPE: internal
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Hepatitis virus

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:55:

CTGGTTCCGC GTGGATCGCA GCGAATTCTC GAGGATCCAT CCCGGCCGCA TCGTGA
56

Claims

We claim:

1. A mosaic protein comprising a plurality of homologous antigenic peptides from different genotypes of a species.
2. The mosaic protein of Claim 1, wherein the species is a hepatitis virus.
3. The mosaic protein of Claim 2, wherein the hepatitis virus is a hepatitis C virus.
4. The mosaic protein of Claim 3, wherein the antigenic peptides are from nucleocapsid proteins.
5. The mosaic protein of Claim 4, comprising the amino acid sequences set forth in SEQ ID NOs:23-33.
6. A method of detecting a hepatitis infection in an individual comprising combining a serum sample from the individual with the mosaic protein of Claim 2, and detecting the presence of antibody binding to the mosaic protein, the presence of binding indicating a hepatitis infection in the individual.
7. The method of Claim 6, wherein the antigenic peptides are from nucleocapsid proteins.
8. The method of Claim 7, wherein the mosaic protein comprises the amino acid sequences set forth in SEQ ID NOs:23-33.
9. The mosaic protein of Claim 3, wherein the antigenic peptides are from non-structural proteins.
10. The mosaic protein of Claim 9, comprising the amino acid sequence set forth in SEQ ID NO:52.

11. The method of Claim 6, wherein the antigenic peptides are from non-structural proteins.

12. The method of Claim 11, wherein the mosaic protein comprises the amino acid sequence set forth in SEQ ID NO:52.

13. A nucleic acid encoding a mosaic protein, wherein the mosaic protein comprises a plurality of homologous antigenic peptides from different genotypes of a species.

14. The nucleic acid of Claim 13, wherein the species is a hepatitis virus.

15. The nucleic acid of Claim 14, wherein the hepatitis virus is a hepatitis C virus.

16. The nucleic acid of Claim 15, wherein the antigenic peptides are from nucleocapsid proteins.

17. The nucleic acid of Claim 16, wherein the mosaic protein comprises the amino acid sequences set forth in SEQ ID NOs:23-33.

18. The nucleic acid of Claim 15, wherein the antigenic peptides are from non-structural proteins.

19. The nucleic acid of Claim 18, wherein the mosaic protein comprises the amino acid sequence set forth in SEQ ID NO:52.

20. A method of constructing an artificial gene, comprising:

a. synthesizing an initial oligonucleotide containing an initial gene segment encoding an initial gene product,

wherein the initial gene segment is flanked in the upstream direction by an upstream initial ligating sequence, a first endonuclease recognition sequence that is recognized by a first endonuclease that cleaves at the first endonuclease recognition sequence, and a second endonuclease recognition sequence which is recognized by a second endonuclease that cleaves downstream of the first endonuclease recognition sequence and within the upstream initial ligating sequence,

and wherein the initial gene segment is flanked in the downstream direction by a downstream initial ligating sequence, a stop codon, a third endonuclease recognition sequence that is recognized by a third endonuclease that cleaves at the third endonuclease recognition sequence, and a fourth endonuclease recognition sequence which is recognized by a fourth endonuclease that cleaves upstream of the third endonuclease recognition sequence, upstream of the stop codon, and within the downstream initial ligating sequence;

b. synthesizing a subsequent oligonucleotide containing a subsequent gene segment encoding a subsequent gene product,

wherein the subsequent gene segment is flanked in the upstream direction by an upstream subsequent ligating sequence, a first endonuclease recognition sequence which is recognized by the first endonuclease that cleaves at the first endonuclease recognition sequence, and a second endonuclease recognition sequence which is recognized by the second endonuclease that cleaves downstream of the first endonuclease recognition sequence and within the upstream subsequent ligating sequence,

and wherein the subsequent gene segment is flanked in the downstream direction by a downstream subsequent ligating sequence, a stop codon, a third endonuclease recognition sequence which is recognized by the third endonuclease that cleaves at the third endonuclease recognition sequence, and a fourth endonuclease recognition sequence which is recognized by

the fourth endonuclease that cleaves upstream of the third endonuclease recognition sequence, upstream of the stop codon, and within the downstream subsequent ligating sequence;

c. cleaving the initial oligonucleotide with the fourth endonuclease, and cleaving the subsequent oligonucleotide with the second endonuclease; and

d. ligating the initial oligonucleotide and the subsequent oligonucleotide together at the downstream initial ligating sequence of the initial oligonucleotide and the upstream subsequent ligating sequence of the subsequent oligonucleotide to form an artificial gene.

21. The method of Claim 20, further comprising the subsequent step of cleaving the artificial gene with the first and third endonucleases and inserting the remaining artificial gene into a vector previously cleaved with the first and third endonucleases.

22. The method of Claim 20, further comprising following the synthesizing the initial oligonucleotide step, and before the cleaving the initial oligonucleotide with the fourth endonuclease step, the additional step of:

confirming the operability of the initial oligonucleotide by cleaving the initial oligonucleotide with the first and third endonucleases and inserting the remaining initial oligonucleotide into a vector previously cleaved with the first and third endonucleases, and expressing the initial gene segment.

23. The method of Claim 20, wherein the first endonuclease is EcoRI, the second endonuclease is BbvI, the third endonuclease is BamHI, or the fourth endonuclease is FokI.

24. The method of Claim 20, wherein the gene segments encode antigenic regions of different genotypes of a hepatitis virus.

25. The method of Claim 20, wherein the gene segments encode antigenic regions of different genotypes of a hepatitis C virus.

26. The method of Claim 20, wherein the gene segments encode antigenic regions of a nucleocapsid protein or a non-structural protein of different genotypes of a hepatitis C virus.

27. An artificial gene constructed by the method of Claim 20.

28. A mosaic protein encoded by the artificial gene constructed by the method of Claim 20.

29. The mosaic protein of Claim 28, wherein the gene segments encode antigenic regions of different genotypes of a hepatitis virus.

30. A method of detecting a hepatitis infection in an individual comprising combining a serum sample from the individual with the mosaic protein of Claim 29, and detecting the presence of antibody binding to the mosaic protein, the presence of binding indicating a hepatitis infection in the individual.

Abstract**MOSAIC PROTEIN AND RESTRICTION
ENDONUCLEASE ASSISTED LIGATION METHOD
FOR MAKING THE SAME**

A mosaic protein comprising a variety of immunoreactive antigenic epitopes from several genotypes of hepatitis C virus. The mosaic protein provides a sensitive and specific immunological hepatitis detection assay. A restriction enzyme assisted ligation method of making an artificial gene permits controlled construction of mosaic proteins, and allows confirmatory expression of the intermediate gene products.

Attorney Docket: 03063-0380

CDC Docket I-018-97/0

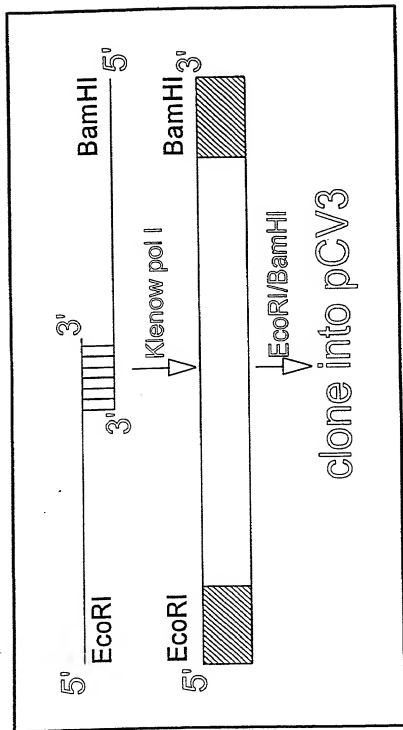


FIG. 1

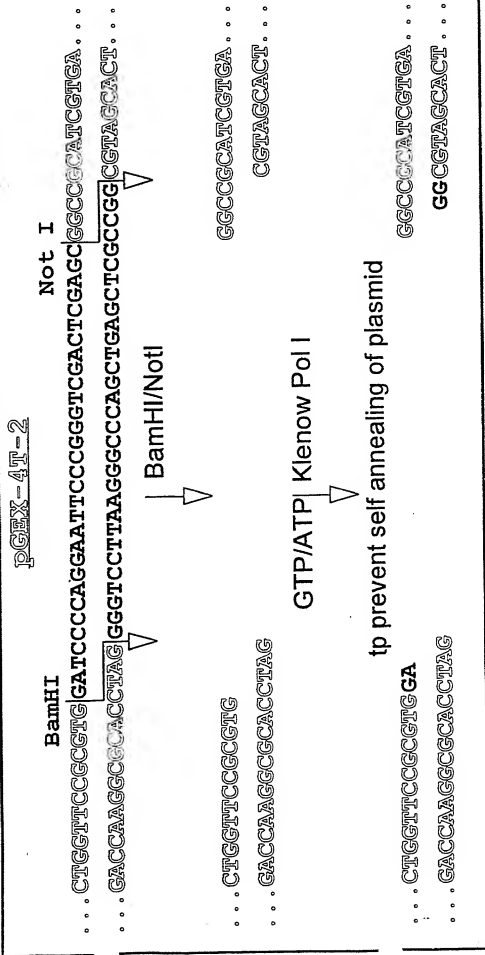


FIG. 2

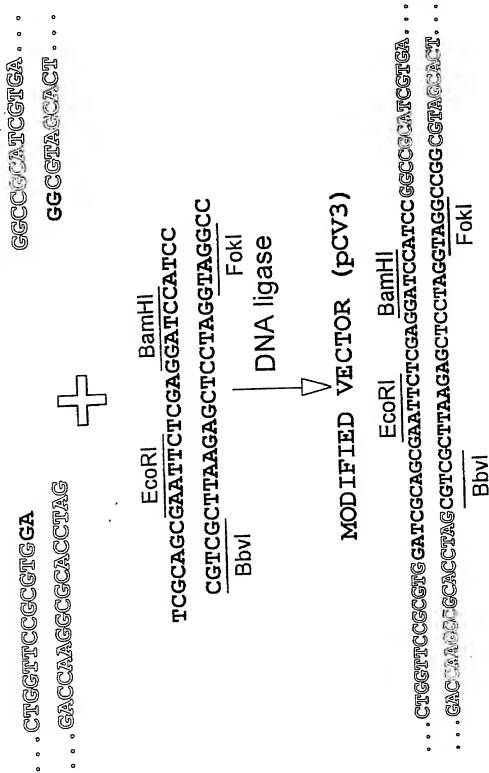


FIG. 3

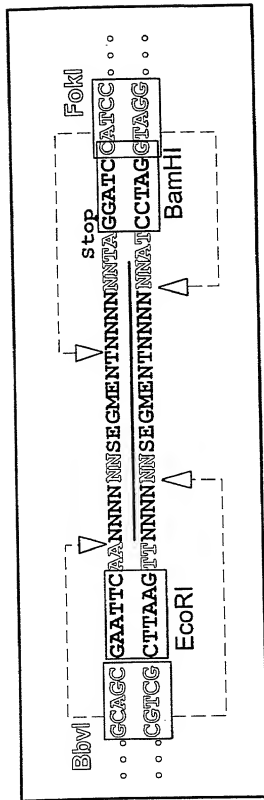


FIG. 4

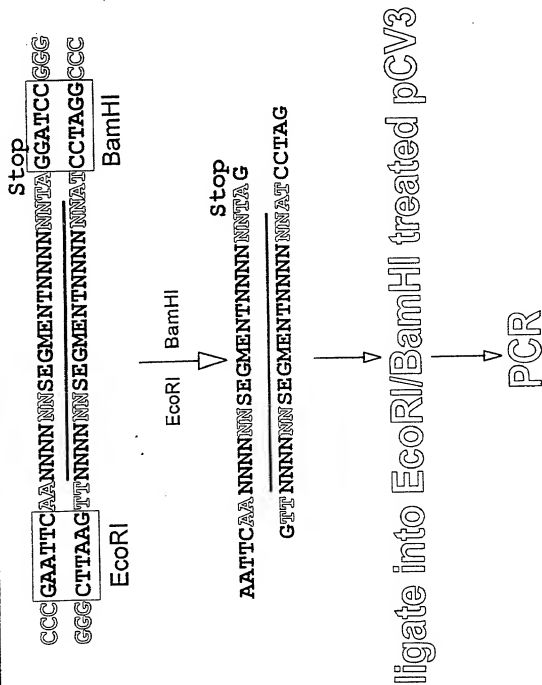


FIG. 5

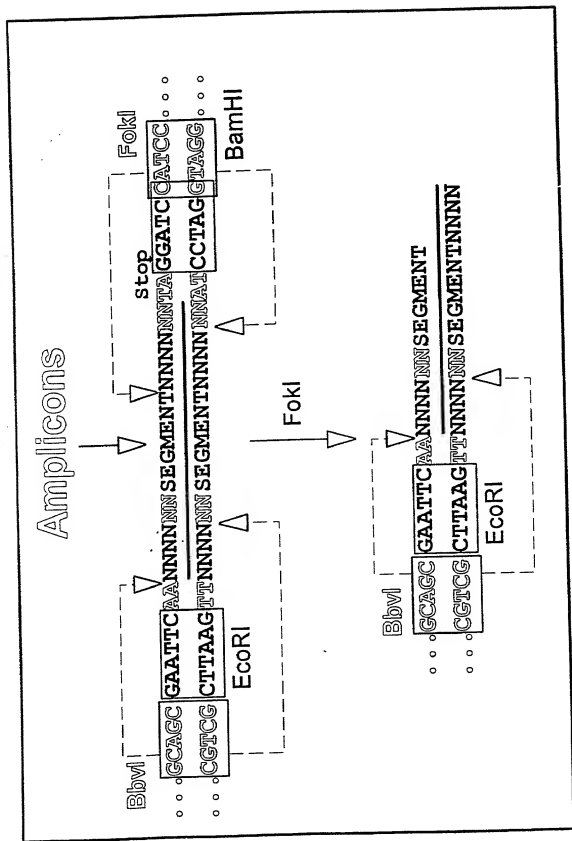


FIG. 6

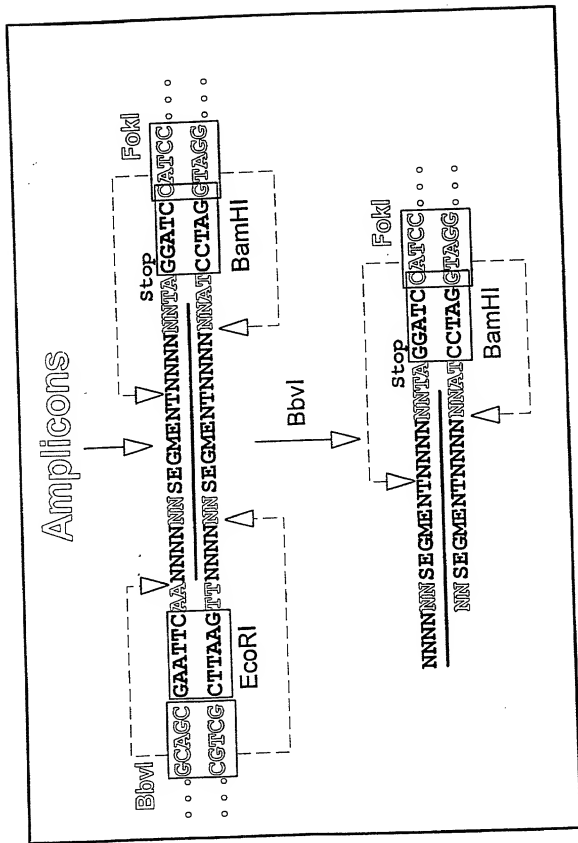


FIG. 7

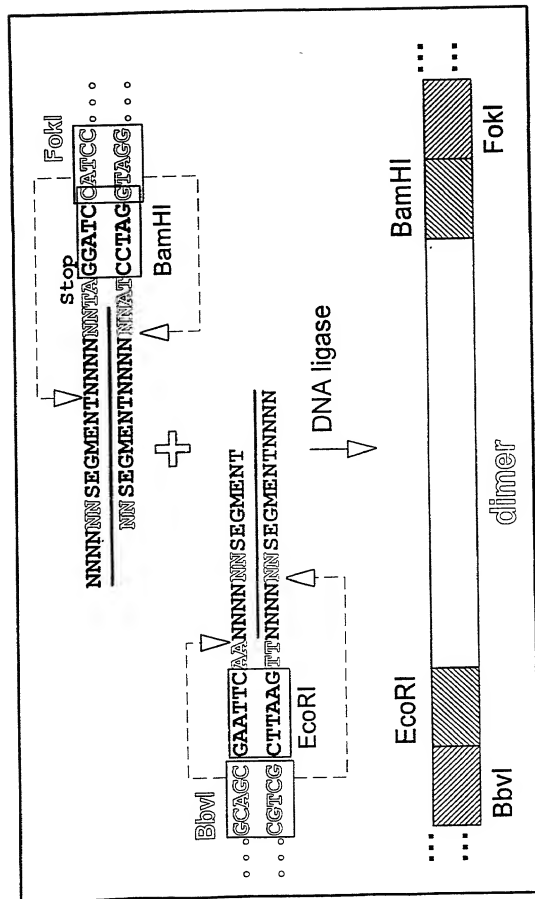
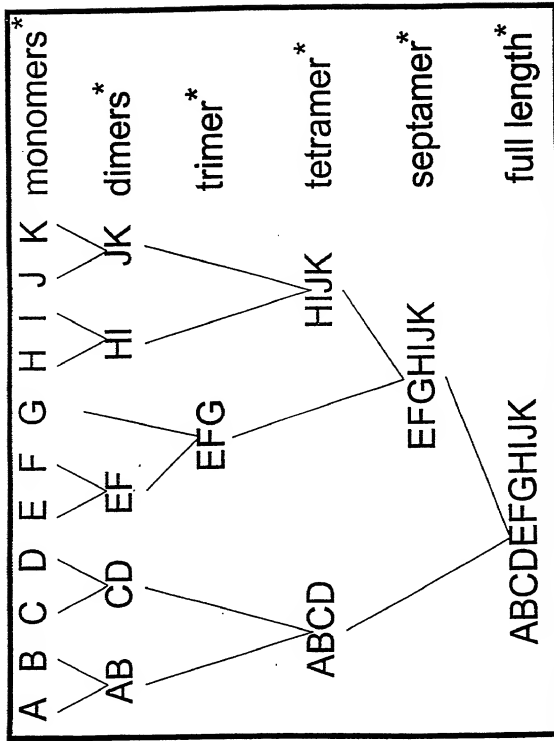


FIG. 8

Fragment	Sequence	SEQ ID NO
A	PKPQRTKRN TIRRPQDVKF PGGQIVG	23
B	PKPQRTKRN TNRRPQDVKF PGGQIVG	24
C	PKPQRTKRN TYRRPQDVKF PGGQIVG	25
D	PKPQRTKRN TNRRPQDVKF PGGQIVG	26
E	PKPQRTKRN TPRRPQDVKF PGGQIVG	27
F	PKPQRTKRN AHRRPQDVKF PGGQIVG	28
G	PKPQRTKRN TNRRPQDVKF PGGQIVG	29
H	PKPQRTKRN TIRRPQDVKF PGGQIVG	30
I	PKPQRTKRN TNRRPQDVKF PGGQIVG	31
J	PKPQRTKRN TIRRPQDVKF PGGQIVG	32
K	PKPQRTKRN AHRRPQDVKF PGGQIVG	33

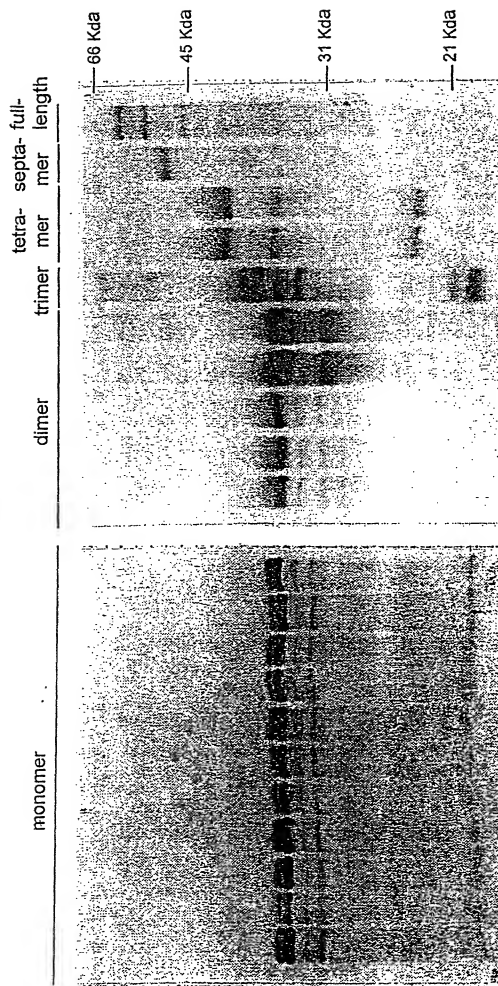
FIG. 9

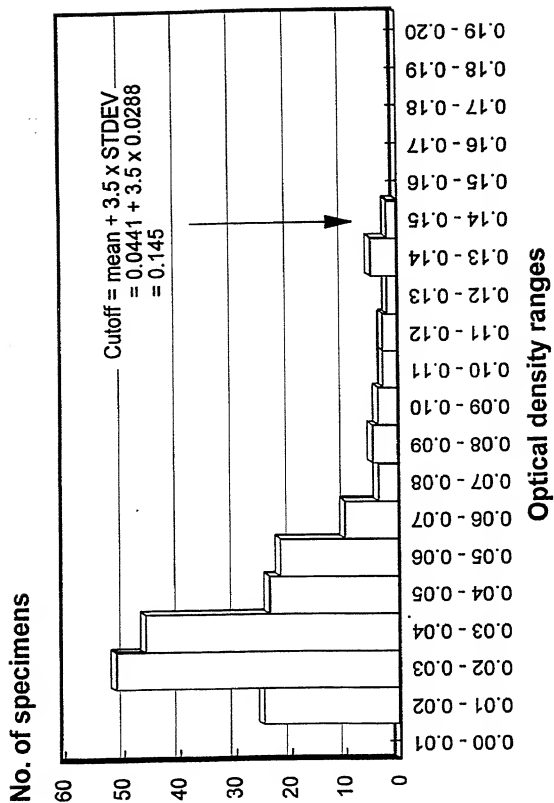


* PCR, clone, PCR for sequence confirmation and sequential assembly, express, determine immunoreactivity

FIG. 10

Figure 11



**FIG. 12**

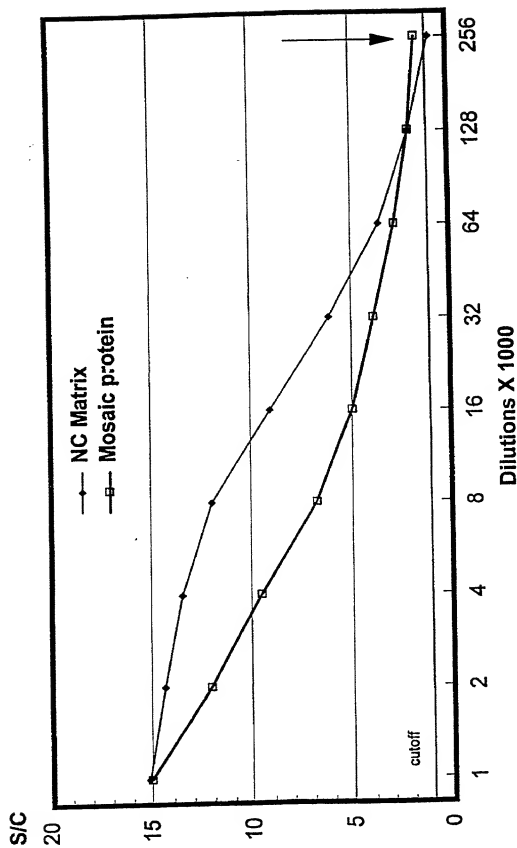
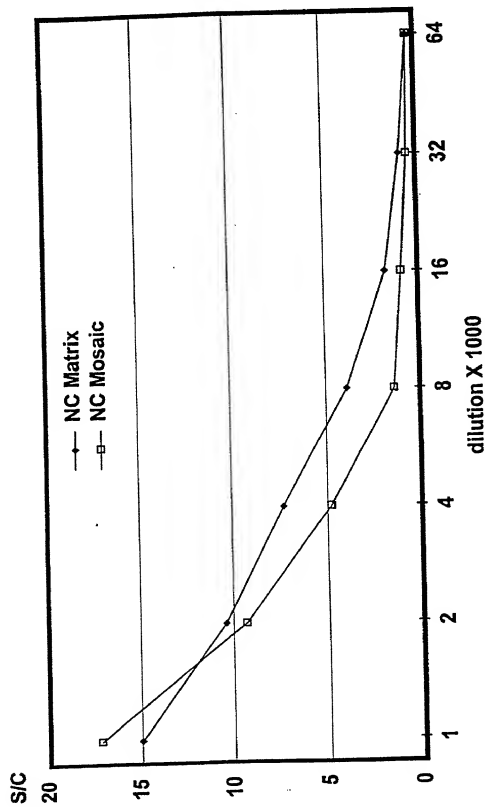


FIG. 13

**FIG. 14**

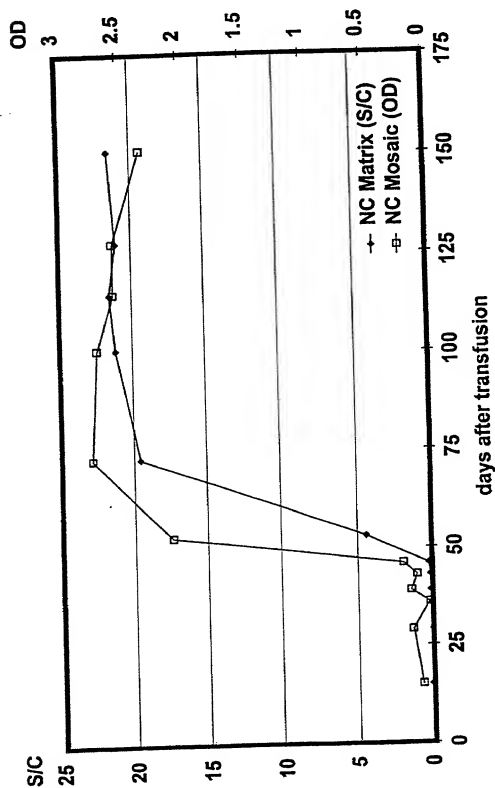


FIG. 15

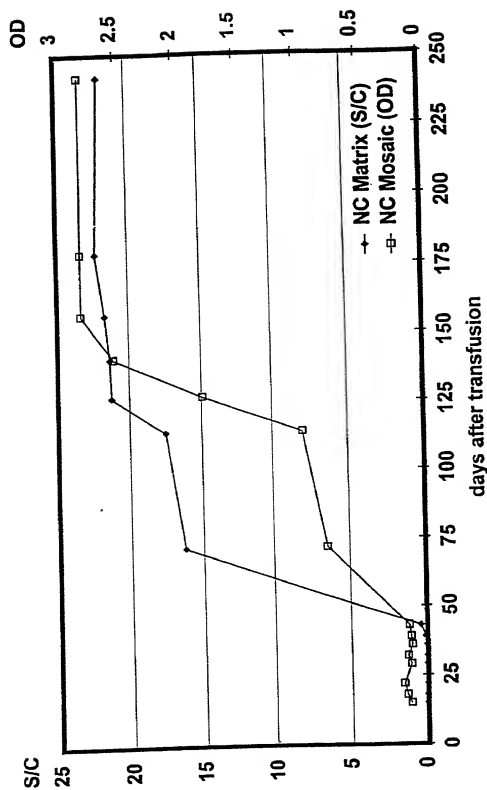


FIG. 16

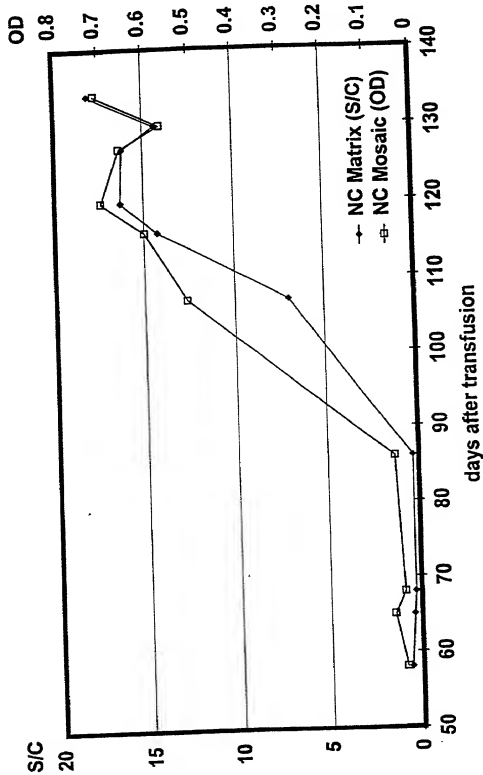


FIG. 17

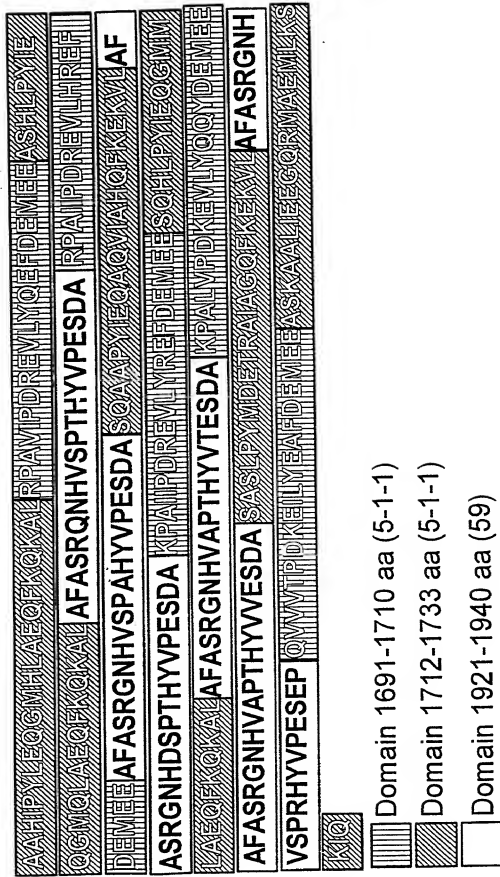
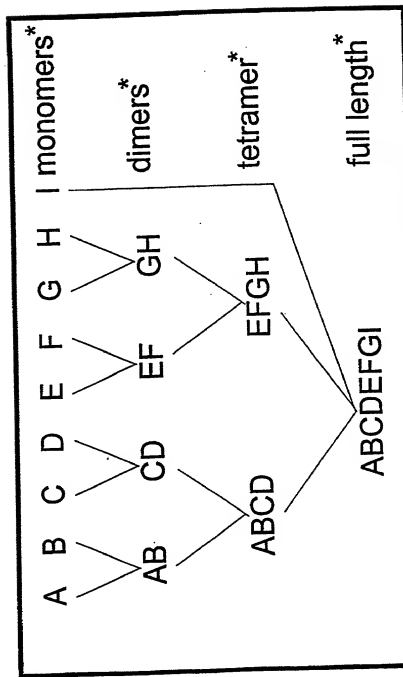


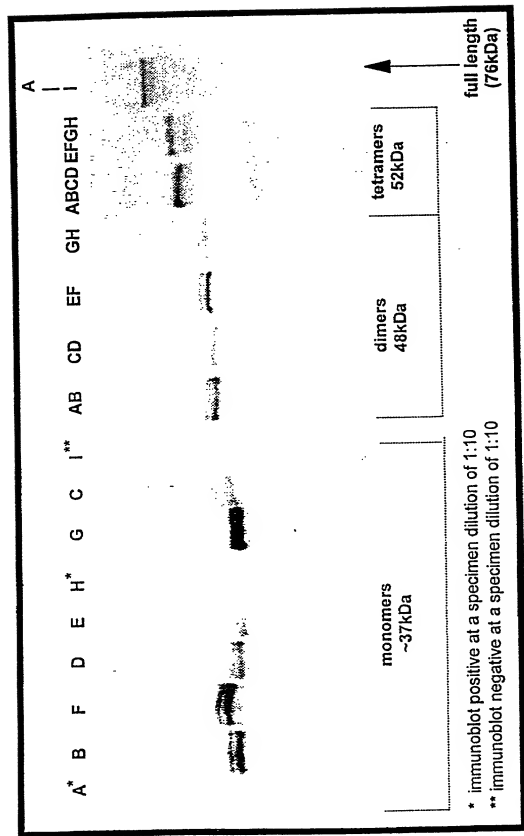
FIG. 18



* PCR, clone, PCR for sequence confirmation and sequential assembly, express, determine immunoreactivity

FIG. 19

Figure 20



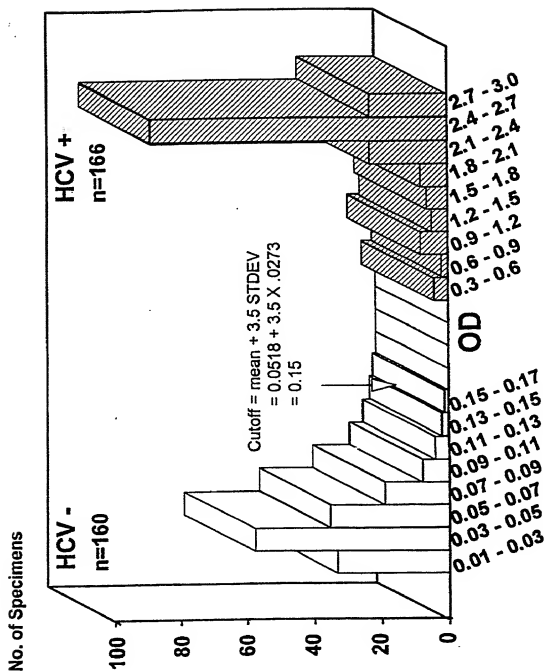


FIG. 21

FIG. 22A

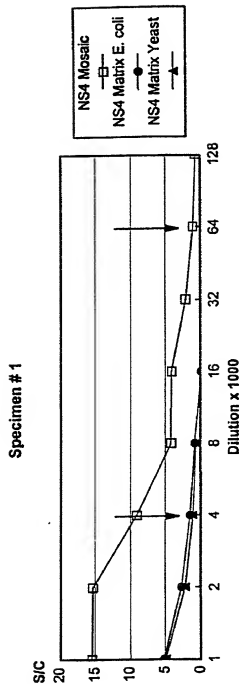
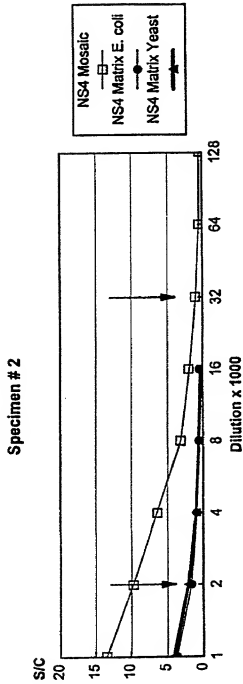


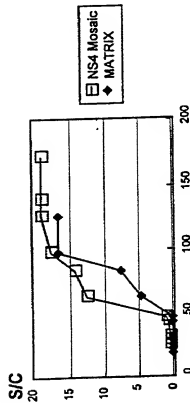
FIG. 22B



	NC Matrix	NS3 Matrix	NS4 Matrix	NS4 Mosaic
No. of anti-HCV sera tested	182	182	182	182
No. of positive sera	172	179	158	178
Percent positive	94.5%	98.4%	86.8%	97.8%

FIG. 23

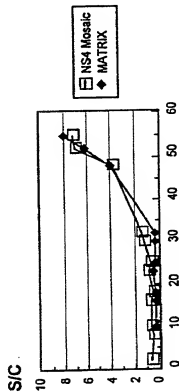
4811



Days After Transfusion

FIG. 24A

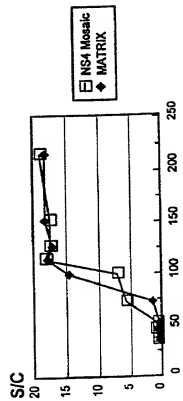
6214



Days After Transfusion

FIG. 24B

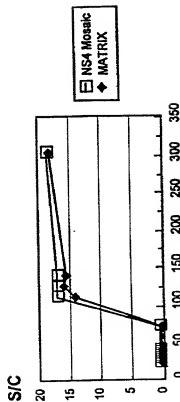
4812



Days After Transfusion

FIG. 24C

4813



Days After Transfusion

FIG. 24D

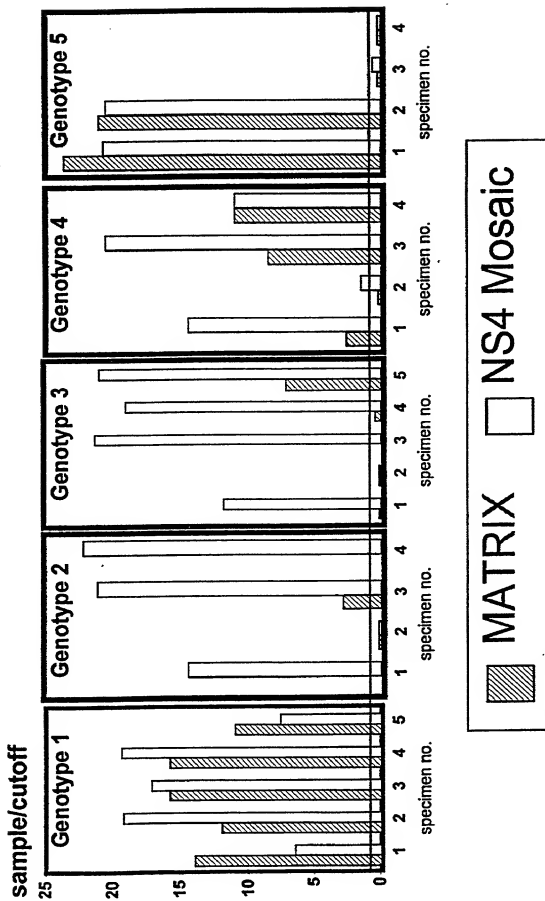


FIG. 25

DECLARATION AND POWER O. ATTORNEYAttorney's Docket No. **03063-0380**
CDC No. I-018-97/0In re Application of: **Yury E. Khudyakov and Howard A. Fields**

As a below named inventor, I hereby declare that:

My residence, post office address and citizenship are as stated below next to my name. I believe I am an original, first and joint inventor (if plural names are listed above) of the subject matter which is claimed and for which a patent is sought on the invention entitled: **MOSAIC PROTEIN AND RESTRICTION ENDONUCLEASE ASSISTED LIGATION METHOD FOR MAKING THE SAME**, the specification of which was filed on **August 25, 1997**, as Application No. **08/921,887**.

I hereby state that I have reviewed and understand the contents of the above-identified specification, including the claims, as amended by any amendment referred to above. I do not know and do not believe that the same was ever known or used by others in the United States of America before my or our invention thereof, or patented or described in any printed publication in any country before my or our invention thereof or more than one year prior to the date of this application. I further state that the invention was not in public use or on sale in the United States of America more than one year prior to the date of this application. *I understand that I have a duty of candor and good faith toward the Patent and Trademark Office, and I acknowledge the duty to disclose information which is material to the examination of this application in accordance with Title 37, Code of Federal Regulations, §1.56.*

I hereby claim foreign priority benefits under Title 35, United States Code, §119 (a)-(d) of the foreign application(s) for patent or inventor's certificate listed below, and have also identified below any foreign application for patent or inventor's certificate disclosing subject matter in common with the above-identified specification and having a filing date before that of the application on which priority is claimed:

<u>Application No.</u>	<u>Country</u>	<u>Filing Date</u>	<u>Priority Claimed Under 35 USC §119</u>
			Yes _____ No _____

I hereby claim the benefit under Title 35, United States Code, § 119(e) of any United States provisional application(s) listed below:

<u>(Application No.)</u>	<u>(Filing Date)</u>	<u>(Application No.)</u>	<u>(Filing Date)</u>
--------------------------	----------------------	--------------------------	----------------------

I hereby claim the benefit under Title 35, United States Code, §120 of any United States application(s) listed below and, insofar as the subject matter disclosed and claimed in the present application is not disclosed in the prior United States application in the manner provided by the first paragraph of Title 35, United States Code §112, I acknowledge the duty to disclose material information as defined in Title 37, Code of Federal Regulations, §1.56 which became available between the filing date of the prior application and the national or PCT international filing date of this application:

<u>Application Serial No.</u>	<u>Filing Date</u>	<u>Status: patented, pending, abandoned</u>
-------------------------------	--------------------	---

I further declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code, and that such willful false statements may jeopardize the validity of the application or any patents issuing thereon.

POWER OF ATTORNEY: The following attorneys are hereby appointed to prosecute this application and transact all business in the Patent and Trademark Office connected therewith: Anthony B. Askew - 24,154; Roger T. Frost - 22,176; Jeffrey E. Young - 28,490; Robert E. Richards - 29,105; John R. Harris - 30,388; Stephen M. Schaezel - 31,418; Larry A. Roberts - 31,871; Thomas A. Hodge - 22,602; Charles L. Warner II - 32,320; Gregory T. Gronholm - 32,415; Dale Lischer - 28,438; Peter G. Pappas - 33,205; James Dean Johnson - 31,771; Nora M. Tocups - 35,717; W. Scott Petty - 35,645; Daniel J. Warren - 34,272; Hubert J. Barnhardt III - 36,739; Virginia L. Carron - 37,110; Leona G. Young - 37,266; Jamie L. Greene-32,467; Holmes J. Hawkins III - 38,913; Mary Anthony Merchant - 39,771; Michael J. Mehrman - 40,086; William L. Warren - 36,714; Felipe J. Farley - 38,445; F. Leslie Bessenger III - 39,108; James A. Witherspoon - 36,723; Brenda M. Ozaki - 40,339; James D. Withers - 40,376; M. Todd Mitchem - P40,731; Gregory S. Smith - P40,819; John J. Snyder - 39,380; Frances L. Reid-Sanden - 38,592; Marjorie D. Hunter - 30,560.

Send correspondence to: **JONES & ASKEW, LLP**
191 Peachtree Street, N.E., 37th Floor
Atlanta, GA 30303-1769

Direct telephone calls at **(404) 818-3700**

William L. Warren

Full name of joint inventor: Yury E. Khudyakov	Citizenship: Russia
Residence and Post Office Address: 4354 Tremont Court, N.W., Duluth, GA 30136	
Inventor's signature: <i>Yury E. Khudyakov</i>	Date: <i>10/14/97</i>

Full name of joint inventor: Howard A. Fields	Citizenship: U.S.A.
Residence and Post Office Address: 1823 Jackson Creek Drive, Marietta, GA 30068	
Inventor's signature: <i>Howard A. Fields</i>	Date: <i>10/13/97</i>